

Hyperbolic Cross Approximation

Dinh Dũng^a, Vladimir Temlyakov^{b*}, Tino Ullrich^c

^aInformation Technology Institute, Vietnam National University, Hanoi, Vietnam

^bDepartment of Mathematics, University of South Carolina, 29208 Columbia, USA

^cInstitute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany

December 5, 2016

Abstract

Hyperbolic cross approximation is a special type of multivariate approximation. Recently, driven by applications in engineering, biology, medicine and other areas of science new challenging problems have appeared. The common feature of these problems is high dimensions. We present here a survey on classical methods developed in multivariate approximation theory, which are known to work very well for moderate dimensions and which have potential for applications in really high dimensions. The theory of hyperbolic cross approximation and related theory of functions with mixed smoothness are under detailed study for more than 50 years. It is now well understood that this theory is important both for theoretical study and for practical applications. It is also understood that both theoretical analysis and construction of practical algorithms are very difficult problems. This explains why many fundamental problems in this area are still unsolved. Only a few survey papers and monographs on the topic are published. This and recently discovered deep connections between the hyperbolic cross approximation (and related sparse grids) and other areas of mathematics such as probability, discrepancy, and numerical integration motivated us to write this survey. We try to put emphases on the development of ideas and methods rather than list all the known results in the area. We formulate many problems, which, to our knowledge, are open problems. We also include some very recent results on the topic, which sometimes highlight new interesting directions of research. We hope that this survey will stimulate further active research in this fascinating and challenging area of approximation theory and numerical analysis.

*Corresponding author, Email: temlyakovusc@gmail.com

Contents

1	Introduction	4
2	Trigonometric polynomials	14
2.1	Univariate polynomials	14
2.2	Multivariate polynomials	17
2.3	Hyperbolic cross polynomials	20
2.4	The Bernstein-Nikol'skii inequalities	22
2.5	Volume estimates	26
2.6	Riesz products and the Small Ball Inequality	28
2.7	Comments and open problems	30
3	Function spaces on \mathbb{T}^d	32
3.1	Spaces of functions with bounded mixed derivative	32
3.2	Spaces of functions with bounded mixed difference	33
3.3	Characterization via Fourier transform	35
3.4	Embeddings	37
4	Linear approximation	39
4.1	Introduction	39
4.2	Approximation by the hyperbolic cross polynomials	41
4.3	The Kolmogorov widths	46
4.4	Orthowidths	53
4.5	The linear widths	56
4.6	Open problems	60
5	Sampling recovery	61
5.1	The univariate de la Vallée Poussin interpolation	64
5.2	Frequency-limited sampling representations - discrete Littlewood-Paley	65
5.3	Sampling on the Smolyak grids	68
5.4	Sampling widths	72
5.5	Time-limited sampling representations-B-splines	75
5.6	Open problems	80
6	Entropy numbers	82
6.1	General notions and inequalities	82
6.2	Entropy numbers for W classes in L_q	85
6.3	Entropy numbers for H and B classes in L_q	87
6.4	Entropy numbers and the Small Ball Problem	89
6.5	Open problems	91
7	Best m-term approximation	92
7.1	Introduction	92
7.2	Orthogonal bases	93
7.3	Some related problems	97
7.4	A comment on Stechkin's lemma	98
7.5	Sparse trigonometric approximation	100
7.6	Open problems	107

8	Numerical integration	108
8.1	The problem setting	108
8.2	Lower bounds	108
8.3	Cubature on Smolyak grids	110
8.4	The Fibonacci cubature formulas	112
8.5	The Frolov cubature formulas	114
8.6	Modifications of Frolov's method	118
8.7	Quasi-Monte Carlo cubature	120
8.8	Discrepancy and numerical integration	123
8.9	Open problems and historical comments	128
9	Related problems	131
9.1	Why classes with mixed smoothness?	131
9.2	Universality	132
9.3	Further generalizations	135
9.4	Direct and inverse theorems	136
9.5	Kolmogorov widths of the intersection of function classes	136
9.6	Further s -numbers	137
9.7	The quasi-Banach situation	140
9.8	Sampling along lattices	143
9.9	Sampling recovery in energy norm	143
9.10	Continuous algorithms in m -term approximation and nonlinear widths	144
10	High-dimensional approximation	146
10.1	Anisotropic mixed smoothness	146
10.2	Explicit constants and preasymptotics	147
10.3	Approximation in the energy norm	150
10.4	ε -dimension and approximation in infinite dimensions	150
11	Appendix	153
11.1	General notation	153
11.2	Inequalities	153
11.3	Duality in L_p spaces	155
11.4	Fourier series of functions in L_p	156

1 Introduction

This paper is a survey on multivariate approximation. The 20th century was a period of transition from univariate problems to multivariate problems in a number of areas of mathematics. For instance, it is a step from Gaussian sums to Weil's sums in number theory, a step from ordinary differential equations to PDEs, a step from univariate trigonometric series to multivariate trigonometric series in harmonic analysis, a step from quadrature formulas to cubature formulas in numerical integration, a step from univariate function classes to multivariate function classes in approximation theory. In many cases this step brought not only new phenomena but also required new techniques to handle the corresponding multivariate problems. In some cases even a formulation of a multivariate problem requires a nontrivial modification of a univariate problem. For instance, the problem of convergence of the multivariate trigonometric series immediately encounters a question of which partial sums we should consider – there is no natural ordering in the multivariate case. In other words: What is a natural multivariate analog of univariate trigonometric polynomials? Answering this question mathematicians studied different generalizations of the univariate trigonometric polynomials: with frequencies from a ball, a cube or, most importantly, a *hyperbolic cross*

$$\Gamma(N) = \left\{ \mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d : \prod_{j=1}^d \max\{|k_j|, 1\} \leq N \right\}.$$

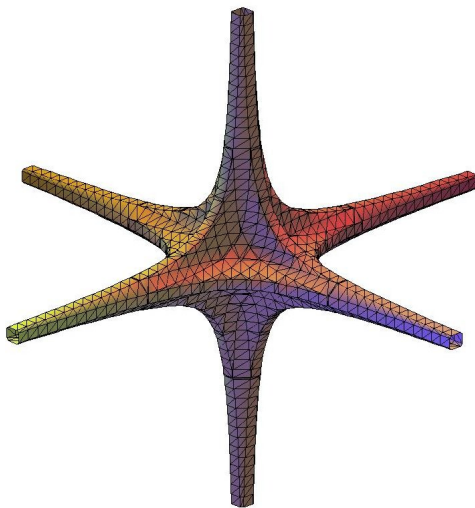


Figure 1: A smooth hyperbolic cross in \mathbb{R}^3

Results discussed in this survey demonstrate that polynomials with frequencies from hyperbolic crosses $\Gamma(N)$ play the same role in the multivariate approximation as the univariate trigonometric polynomials play in the approximation of functions on a single variable. On a very simple example we show how the hyperbolic cross polynomials appear naturally in the multivariate approximation. Let us begin with the univariate case. The natural ordering of the univariate trigonometric system is closely connected with the ordering of eigenvalues of the differential operator $D := \frac{d}{dx}$ considered on 2π periodic functions. The eigenvalues are $\{\pm ik\}$ with $k = 0, 1, 2, \dots$ and the corresponding eigenfunctions are $\{e^{\pm ikx}\}$. Nonzero eigenvalues of the differential operator of mixed derivative $\prod_{j=1}^d D_j$ are $\{\prod_{j=1}^d ik_j : \mathbf{k} \in \mathbb{Z}^d\}$. Ordering these eigenvalues we immediately obtain the hyperbolic crosses $\Gamma(N)$. This simple observation shows

that hyperbolic crosses are closely connected with the mixed derivative. Results obtained in the multivariate approximation theory for the last 50 years established a deep connection between trigonometric polynomials with frequencies from the hyperbolic crosses and classes of functions defined with the help of either mixed derivatives or mixed differences. The importance of these classes was understood in the beginning of 1960s.

In the 1930s in connection with applications in mathematical physics, S.L. Sobolev introduced the classes of functions by imposing the following restrictions

$$\|f^{(n_1, \dots, n_d)}\|_p \leq 1 \quad (1.1)$$

for all $\mathbf{n} = (n_1, \dots, n_d)$ such that $n_1 + \dots + n_d \leq R$. These classes appeared as natural ways to measure smoothness in many multivariate problems including numerical integration. It was established that for Sobolev classes the optimal error of numerical integration by formulas with m nodes is of order $m^{-R/d}$. On the other hand, at the end of 1950s, N.M. Korobov discovered the following phenomenon: Let us consider the class of functions which satisfy (1.1) for all \mathbf{n} such that $n_j \leq r$, $j = 1, \dots, d$ (Korobov considered different, more general classes, but for illustration purposes it is convenient for us to deal with these classes here). Obviously this new class (class of functions with bounded mixed derivative) is much wider than the Sobolev class with $R = rd$. For example, all functions of the form

$$f(x) = \prod_{j=1}^d f_j(x_j), \quad \|f_j^{(r)}\|_p \leq 1,$$

belong to this class, while not necessarily to the Sobolev class (it would require, roughly, $\|f_j^{(rd)}\|_p \leq 1$). Korobov constructed a cubature formula with m nodes which guaranteed the accuracy of numerical integration for this class of order $m^{-r}(\log m)^{rd}$, i.e., almost the same accuracy that we had for the Sobolev class. Korobov's discovery pointed out the importance of the classes of functions with bounded mixed derivative in fields such as approximation theory and numerical analysis. The simplest versions of Korobov's magic cubature formulas are the Fibonacci cubature formulas, see Subsection 8.4 below, given by

$$\Phi_n(f) = b_n^{-1} \sum_{\mu=1}^{b_n} f(\mu/b_n, \{\mu b_{n-1}/b_n\}),$$

where $b_0 = b_1 = 1$, $b_n = b_{n-1} + b_{n-2}$ are the Fibonacci numbers and $\{x\}$ is the fractional part of the number x . These cubature formulas work optimally both for classes of functions with bounded mixed derivative and for classes with bounded mixed difference. The reason for such an outstanding behavior is the fact that the Fibonacci cubature formulas are exact on the hyperbolic cross polynomials associated with $\Gamma(cb_n)$.

The multivariate problems of hyperbolic cross approximation turn out to be much more involved than their univariate counterparts. For instance, the fundamental Bernstein inequalities for the trigonometric polynomials are known in the univariate case with explicit constants and they are not even known in the sense of order for the trigonometric polynomials with frequencies from a hyperbolic cross (see Open problem 1.1 below).

We give a brief historical overview of challenges and open problems of approximation theory with emphasis put on multivariate approximation. It was understood in the beginning of the 20th century that smoothness properties of a univariate function determine the rate of approximation of this function by polynomials (trigonometric in the periodic case and algebraic in the non-periodic case). A fundamental question is: What is a natural multivariate analog

of univariate smoothness classes? Different function classes were considered in the multivariate case: isotropic and anisotropic Sobolev and Besov classes, classes of functions with bounded mixed derivative and others. The simplest case of such a function class is the unit ball of the mixed Sobolev space of bivariate functions given by

$$\mathbf{W}_p^r := \left\{ f \in L_p : \|f\|_{\mathbf{W}_p^r} := \|f\|_p + \left\| \frac{\partial^r f}{\partial x_1^r} \right\|_p + \left\| \frac{\partial^r f}{\partial x_2^r} \right\|_p + \left\| \frac{\partial^{2r} f}{\partial x_1^r \partial x_2^r} \right\|_p \leq 1 \right\}.$$

These classes are sometimes denoted as classes of functions with dominating mixed derivative since the condition on the mixed derivative is the dominating one. Babenko [7] was the first who introduced such classes and began to study approximation of these classes by the hyperbolic cross polynomials. In Section 3 we will define more general periodic \mathbf{W} and \mathbf{H} , \mathbf{B} classes of d -variate functions, also with fractional smoothness $r > 0$. What concerns \mathbf{H} and \mathbf{B} classes we replace the condition on the mixed derivative, used for the definition of \mathbf{W} classes, by a condition on a mixed difference. In Section 3 we give a historical comment on the further study of the mixed smoothness classes.

The next fundamental question is: How to approximate functions from these classes? Kolmogorov introduced the concept of the n -width of a function class. This concept is very useful in answering the above question. The Kolmogorov n -width is a solution to an optimization problem where we minimize the error of best approximation with respect to all n -dimensional linear subspaces. This concept allows us to understand which n -dimensional linear subspace is the best for approximating a given class of functions. The rates of decay of the Kolmogorov n -width are known for the univariate smoothness classes. In some cases even exact values of it are known. The problem of the rates of decay of the Kolmogorov n -width for the classes of multivariate functions with bounded mixed derivative is still not completely understood.

We note that the function classes with bounded mixed derivative are not only an interesting and challenging object for approximation theory. They also represent a suitable model in scientific computations. Bungartz and Griebel [45, 44, 151] and their groups use approximation methods designed for these classes in elliptic variational problems. The recent work of Yserentant [406] on the regularity of eigenfunctions of the electronic Schrödinger operator, and Triebel [384] on the regularity of solutions of Navier-Stokes equations, show that mixed regularity plays a fundamental role in mathematical physics. This makes approximation techniques developed for classes of functions with bounded mixed derivative a proper choice for the numerical treatment of those problems.

Approximation of classes of functions with bounded mixed derivative (the \mathbf{W} classes) and functions with a restriction on mixed differences (the \mathbf{H} and, more generally, \mathbf{B} classes) have been developed following classical tradition. A systematic study of different asymptotic characteristics of these classes dates back to the beginning of 1960s. Babenko [6] and Mityagin [233] were the first who established the nowadays well-known classical estimates for the Kolmogorov widths of the classes \mathbf{W}_p^r in L_p if $1 < p < \infty$, namely

$$d_m(\mathbf{W}_p^r, L_p) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r, \quad m \in \mathbb{N}, \quad (1.2)$$

where the constants behind \asymp depend on d , p and r . Later it turned out that the same order is present in the situation $d_m(\mathbf{W}_p^r, L_q)$ if $1 < q \leq p < \infty$ and $2 \leq p \leq q < \infty$. With (1.2) it has been realized that the *hyperbolic cross polynomials* play the same role in the approximation of multivariate functions as classical univariate polynomials for the approximation of univariate functions. This discovery resulted in a detailed study of the hyperbolic cross polynomials. However, it turned out that the study of properties of hyperbolic cross polynomials is much

more difficult than the study of their univariate analogs. We discuss the corresponding results in Section 2.

In Sections 4 and 5 we consider linear approximation problems – the problems of approximation by elements of a given finite dimensional linear subspace. We discussed a number of the most important asymptotic characteristics related to the linear approximation. In addition to the Kolmogorov n -width we also study the asymptotic behavior of linear widths and orthowidths. Interesting effects occur when studying the approximation of the class \mathbf{W}_p^r in L_q if $p < q$. In contrast to (1.2) the influence of the parameters p and q is always visible in the rate of the order of the linear widths. In fact, if either $1 < p \leq q \leq 2$ or $2 \leq p \leq q < \infty$ then we have

$$\lambda_m(\mathbf{W}_p^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/q}, \quad m \in \mathbb{N}.$$

Then the optimal approximant is realized by a projection on an appropriate linear subspace of the hyperbolic cross polynomials, which is not always the case, like for instance in the case $p < 2 < q$. In the definition of linear width we allow all linear operators of rank m to compete in the minimization problem. Clearly, we would like to work with nice and simple linear operators. This idea motivated researchers to impose additional restrictions on the linear operators in the definitions of the corresponding modifications of the linear width. One very natural restriction is that the approximating rank m operator is an orthogonal projection operator. This leads to the concept of the orthowidth (Fourier width) $\varphi_m(\mathbf{F}, X)$. It turns out that the behavior of the $\varphi_m(\mathbf{W}_p^r, L_q)$ is different from the behavior of the $\lambda_m(\mathbf{W}_p^r, L_q)$. For instance, it was proved that the operators S_{Q_n} of orthogonal projection onto subspaces $\mathcal{T}(Q_n)$ of the hyperbolic cross polynomials are optimal (in the sense of order) from the point of view of the orthowidth for all $1 \leq p, q \leq \infty$ except $p = q = 1$ and $p = q = \infty$. That proof required new nontrivial methods for establishing the right lower bounds.

Another natural restriction is that the approximating rank m operator is a recovering operator, which uses function values at m points. Restricting the set of admissible rank m operators to such that are based on function evaluations (instead of general linear functionals), we observe a behavior which is clearly bounded below by λ_m . We call the corresponding asymptotic quantities sampling widths ϱ_m . However, this is not the end of the story. Already in the situation \mathbf{W}_p^r in L_q we are able to determine sets of parameters where ϱ_m is equal to λ_m in the sense of order, and others where ϱ_m behaves strictly worse (already in the main rate). However, the complete picture is still unknown. In that sense the situation $p = q$ (including $p = 2$) is of particular interest. The result

$$\varrho_m(\mathbf{W}_p^r, L_p) \lesssim \left(\frac{(\log m)^{(d-1)}}{m} \right)^r (\log m)^{(d-1)/2}, \quad m \in \mathbb{N},$$

has been a breakthrough since it improved on a standard upper bound by using a non-trivial technique. However, the exact order is still unknown even in case $p = 2$. The so far best-known upper bounds for sampling recovery are all based on sparse grid constructions. Alike the orthowidth results, the optimal in the sense of order subspaces for recovering are subspaces $\mathcal{T}(Q_n)$ with appropriate n (in all cases, where we know the order of $\varrho_m(\mathbf{W}_p^r, L_q)$).

In Section 6 we discuss a very important characteristic of a compact – its entropy numbers. It quantitatively determines its “degree of compactness”. Contrary to the asymptotic characteristics discussed in Sections 4 and 5 the entropy numbers are not directly connected to the linear theory of approximation. However, there are very useful inequalities between the entropy numbers and other asymptotic characteristics, which provide a powerful method of proving good lower bounds, say, for the Kolmogorov widths. In this case Carl’s inequality is used. Two more

points, which motivated us to discuss the entropy numbers of classes of functions with bounded mixed derivative are the following. (A) The problem of the rate of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$ in a particular case $r = 1, p = 2, q = \infty$ is equivalent to a fundamental problem of probability theory (the Small Ball Problem, see Subsection 6.4 for details). Both of these problems are still open for $d > 2$. (B) The problem of the rate of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$ turns out to be a very rich and difficult problem, which is not yet completely solved. Those problems that have been resolved required different nontrivial methods for different pairs (p, q) .

Here is a typical result on the $\epsilon_n(\mathbf{W}_p^r, L_q)$: For $1 < p, q < \infty$ and $r > (1/p - 1/q)_+$ one has

$$\epsilon_n(\mathbf{W}_p^r, L_q) \asymp \left(\frac{(\log n)^{d-1}}{n} \right)^r, \quad n \in \mathbb{N}.$$

The above rate of decay does neither depend on p nor on q . It is known in approximation theory that investigation of asymptotic characteristics of classes \mathbf{W}_p^r in L_q becomes more difficult when p or q takes value 1 or ∞ than when $1 < p, q < \infty$. This is true for $\epsilon_n(\mathbf{W}_p^r, L_q)$, too. There are still fundamental open problems, for instance, Open problem 1.6 below.

Recently, driven by applications in engineering, biology, medicine and other areas of science nonlinear approximation began to play an important role. Nonlinear approximation is important in applications because of its concise representations and increased computational efficiency. In Section 7 we discuss a typical problem of nonlinear approximation – the m -term approximation. Another name for m -term approximation is *sparse approximation*. In this setting we begin with a given system of elements (functions) \mathcal{D} , which is usually called a *dictionary*, in a Banach space X . The following characteristic

$$\sigma_m(f, \mathcal{D})_X := \inf_{\substack{g_i \in \mathcal{D}, c_i \\ i=1, \dots, m}} \left\| f - \sum_{i=1}^m c_i g_i \right\|_X$$

is called best m -term approximation of f with regard to \mathcal{D} and gives us the bottom line of m -term approximation of f . For instance, we can use the classical trigonometric system \mathcal{T}^d as a dictionary \mathcal{D} . Then, clearly, for any $f \in L_p$ we have $\sigma_m(f, \mathcal{T}^d)_{L_p} \leq E_{Q_n}(f)_p$, when $m = |Q_n|$ (see Section 2 for the definition of the step hyperbolic cross Q_n). Here, $E_{Q_n}(f)_p$ is the best approximation by the hyperbolic cross polynomials with frequencies from the hyperbolic cross Q_n . It turns out that for some function classes best m -term approximations give the same order of approximation as the corresponding hyperbolic cross polynomials but for other classes the nonlinear way of approximation provides a substantial gain over the hyperbolic cross approximation. For instance, when $m = |Q_n|$, we have

$$\sup_{f \in \mathbf{W}_p^r} \sigma_m(f, \mathcal{T}^d)_{L_p} \asymp \sup_{f \in \mathbf{W}_p^r} E_{Q_n}(f)_p, \quad 1 < p < \infty,$$

but for $2 \leq p < q < \infty$ the quantity $\sup_{f \in \mathbf{W}_p^r} \sigma_m(f, \mathcal{D})_{L_q}$ is substantially smaller than $\sup_{f \in \mathbf{W}_p^r} E_{Q_n}(f)_q$.

In a way similar to optimization over linear subspaces in the case of linear approximation we discuss an optimization over dictionaries from a given collection in the m -term approximation problem. It turns out that the wavelet type bases, for instance, the basis \mathcal{U}^d discussed in Section 7, are very good for sparse approximation – in many cases they are optimal (in the sense of order) among all orthogonal bases. A typical result here is the following: for $1 < p, q < \infty$ and large enough r we have

$$\sup_{f \in \mathbf{W}_p^r} \sigma_m(f, \mathcal{U}^d)_{L_q} \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r, \quad m \in \mathbb{N}.$$

It is important to point out that the above rate of decay does not depend on p and q .

The characteristic $\sigma_m(f, \mathcal{D})_X$ gives us a bench mark, which we can ideally achieve in m -term approximation of f . Clearly, keeping in mind possible numerical applications, we would like to devise good constructive methods (algorithms) of m -term approximation. It turns out that greedy approximations (algorithms) work very well for a wide variety of dictionaries \mathcal{D} (see Section 7 for more details).

Numerical integration, discussed in Section 8, is one more challenging multivariate problem where approximation theory methods are very useful. In the simplest form (Quasi-Monte Carlo setting), for a given function class \mathbf{F} we want to find m points $\mathbf{x}^1, \dots, \mathbf{x}^m$ in D such that $\sum_{j=1}^m \frac{1}{m} f(\mathbf{x}^j)$ approximates well the integral $\int_D f d\mu$, where μ is the normalized Lebesgue measure on D . Classical discrepancy theory provides constructions of point sets that are good for numerical integration of characteristic functions of parallelepipeds of the form $P = \prod_{j=1}^d [a_j, b_j]$. The typical error bound is of the form $m^{-1}(\log m)^{d-1}$, see Subsection 8.8 below. Note that a regular grid for $m = n^d$ provides an error of the order $m^{-1/d}$. The above mentioned results of discrepancy theory are closely related to numerical integration of functions with bounded mixed derivative (the case of the first mixed derivative) by the Koksma-Hlawka inequality.

In a somewhat more general setting (optimal cubature formula setting) we are optimizing not only over points $\mathbf{x}^1, \dots, \mathbf{x}^m$ but also over the weights $\lambda_1, \dots, \lambda_m$:

$$\kappa_m(\mathbf{F}) := \inf_{X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\} \subset D} \inf_{\lambda_1, \dots, \lambda_m \in \mathbb{R}} \sup_{f \in \mathbf{F}} \left| \int_D f(\mathbf{x}) d\mathbf{x} - \sum_{i=1}^m \lambda_i f(\mathbf{x}^i) \right|.$$

A typical and very nontrivial result here is: For $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$ we have

$$\kappa_m(\mathbf{W}_p^r) \asymp m^{-r}(\log m)^{(d-1)/2}, \quad m \in \mathbb{N}.$$

In the case of functions of two variables optimal cubature rules are very simple – the Fibonacci cubature rules. They represent a special type of cubature rules, so-called Quasi-Monte Carlo rules. In the case $d \geq 3$ the optimal (in the sense of order) cubature rules are constructive but not as simple as the Fibonacci cubature formulas – the Frolov cubature formulae. In fact, for the Frolov cubature formulae all weights λ_i , $i = 1, \dots, m$, are equal but in general do not sum up to one. This means in particular that constant functions would not be integrated exactly by Frolov’s method. Equal weights which sum up to one is the main feature of the quasi-Monte Carlo integration. We point out that there are still fundamental open problems: right orders of $\kappa_m(\mathbf{W}_p^r)$ for $p = 1$ and $p = \infty$ are not known (see Open problems 1.8 and 1.9 below).

In this survey the notion *Sparse Grids* actually refers to the point grid coming out of Smolyak’s algorithm applied to univariate interpolation/cubature rules. The phrase itself is due to Zenger [407], [154] and co-workers who addressed more general hierarchical methods for avoiding the “full grid” decomposition. In any sense *sparse grids* play an important role in numerical integration of functions with bounded mixed derivative. In the case of r th bounded mixed derivatives they provide an error of the order $m^{-r}(\log m)^{(d-1)(r+1/2)}$. Also, they provide the recovery error in the sampling problem of the same order. Note again that the regular grid from above provides an error of the order $m^{-r/d}$. The error bound $m^{-r}(\log m)^{(d-1)(r+1/2)}$ is reasonably good for moderate dimensions d , say, $d \leq 40$. It turns out that there are practical computational problems with moderate dimensions where sparse grids work well. Sparse grids techniques have applications in quantum mechanics, numerical solutions of stochastic PDEs, data mining, finance.

At the end of this section we give some remarks, which demonstrate a typical difficulty of the study of the hyperbolic cross approximations. It is known that the Dirichlet and the

de la Vallée Poussin kernels and operators, associated with them, play a significant role in investigation of the trigonometric approximation. In particular, the boundedness property of the de la Vallée Poussin kernel: $\|\mathcal{V}_n(x)\|_1 \leq C$ is very helpful. It turns out that there is no analog of this boundedness property for the hyperbolic cross polynomials: for the corresponding de la Vallée Poussin kernel we have $\|\mathcal{V}_{Q_n}(\mathbf{x})\|_1 \geq cn^{d-1}$. This phenomenon made the study in the L_1 and L_∞ norms difficult. There are many unsolved problems for approximation in the L_1 and L_∞ norms.

In the case of L_p , $1 < p < \infty$, the classical tools of harmonic analysis – the Littlewood-Paley theorem, the Marcinkiewicz multipliers, the Hardy-Littlewood inequality – are very useful. However, in some cases other methods were needed to obtain correct estimates. We illustrate it on the following example. Let $\rho(\mathbf{s}) := \{\mathbf{k} : [2^{s_j-1}] \leq |k_j| < 2^{s_j}, j = 1, \dots, d\}$ and

$$\mathcal{D}_{Q_n}(\mathbf{x}) := \sum_{|\mathbf{s}|_1 \leq n} \mathcal{D}_{\rho(\mathbf{s})}(\mathbf{x}), \quad \mathcal{D}_{\rho(\mathbf{s})}(\mathbf{x}) := \sum_{\mathbf{k} \in \rho(\mathbf{s})} e^{i(\mathbf{k}, \mathbf{x})},$$

be the Dirichlet kernel for the step hyperbolic cross Q_n . Then for $2 < p < \infty$ by a corollary to the Littlewood-Paley theorem one gets

$$\|\mathcal{D}_{Q_n}\|_p \lesssim \left(\sum_{|\mathbf{s}|_1 \leq n} \|\mathcal{D}_{\rho(\mathbf{s})}\|_p^2 \right)^{1/2} \lesssim 2^{(1-1/p)n} n^{(d-1)/2}. \quad (1.3)$$

However, the upper bound in (1.3) does not provide the right bound. Other technique (see Theorem 2.11) gives

$$\|\mathcal{D}_{Q_n}\|_p \lesssim \left(\sum_{|\mathbf{s}|_1 \leq n} \left(\|\mathcal{D}_{\rho(\mathbf{s})}\|_2 2^{|\mathbf{s}|_1(1/2-1/p)} \right)^p \right)^{1/p} \lesssim 2^{(1-1/p)n} n^{(d-1)/p}.$$

The above example demonstrates the problem, which is related to the fact that in the multivariate Littlewood-Paley formula we have many ($\asymp n^{d-1}$) dyadic blocks of the same size (2^n).

It is known that in studying asymptotic characteristics of function classes the discretization technique is useful. Classically, the Marcinkiewicz theorem served as a powerful tool for discretizing the L_p -norm of a trigonometric polynomial. Unfortunately, there is no analog of Marcinkiewicz' theorem for hyperbolic cross polynomials (see Theorem 2.25). An important new technique, which was developed to overcome the above difficulties, is based on volume estimates of the sets of Fourier coefficients of unit balls of trigonometric polynomials (see Subsection 2.5). Later, also within the “wavelet revolution”, sequence space isomorphisms were used to discretize function spaces, see Section 5 for the use of the tensorized Faber-Schauder system.

A standard technique of proving lower bounds for asymptotic characteristics of function classes $(d_m, \lambda_m, \varphi_m, \varrho_m, \epsilon_n, \sigma_m, \kappa_m)$ is based on searching for “bad” functions in an appropriate subspaces of the trigonometric polynomials. Afterwards, using the de la Vallée Poussin operator, we reduce the problem of approximation by arbitrary functions to the problem of approximation by the trigonometric polynomials. The uniform boundedness property of the de la Vallée Poussin operators is fundamentally important in this technique. As we pointed out above the de la Vallée Poussin operators for the hyperbolic cross are not uniformly bounded as operators from L_1 to L_1 and from L_∞ to L_∞ – their norms grow with N as $(\log N)^{d-1}$. In some cases we are able to overcome this difficulty by considering the entropy numbers in the respective

situations. In fact, we use some general inequalities (Carl's inequality, see Section 6), which provide lower bounds for the Kolmogorov widths in terms of the entropy numbers. However, there are still outstanding open problems on the behavior of the entropy numbers in L_∞ (see Subsection 6.1). We also point out that classical techniques, based on Riesz products, turned out to be useful for the hyperbolic cross polynomials.

As we already pointed out we present a survey of results on multivariate approximation – the hyperbolic cross approximation. These results provide a natural generalization of the classical univariate approximation theory results to the case of multivariate functions. We give detailed historical comments only on the multivariate results. Typically, the corresponding univariate results are well-known and could be found in a book on approximation theory, for instance, [261], [73], [220], [352]. We discuss here two types of mixed smoothness classes: (I) the **W**-type classes, which are defined by a restriction on the mixed derivatives (more generally, fractional mixed derivatives); (II) the **H**-type and the **B**-type classes, which are defined by restrictions on the mixed differences. It has become standard in the theory of function spaces to call classes defined by restrictions on derivatives Sobolev or Sobolev-type classes. We follow this tradition in our survey. However, we point out that S.L. Sobolev did not study the classes of functions with bounded mixed derivative. The **B**-type classes are usually called Besov or Besov-type classes. We follow this tradition in our survey too. The **H**-type classes are a special case of the **B**-type classes, namely $\mathbf{H}_p^r = \mathbf{B}_{p,\infty}^r$. Historically, the first investigations were conducted on the **W**-type and **H**-type classes. An interesting phenomenon was discovered. It was established that the behavior of the asymptotic characteristics of classes \mathbf{W}_p^r and \mathbf{H}_p^r measured in L_q are different for $1 < p, q < \infty$. Typically, for $d \geq 2$

$$a_m(\mathbf{W}_p^r, L_q) = o(a_m(\mathbf{H}_p^r, L_q)),$$

where a_m stands for best approximations E_{Q_n} , the Kolmogorov width d_m , the linear width λ_m , the orthonormal width φ_m , the entropy numbers ε_m , and the best m -term approximation σ_m . This shows that classes \mathbf{H}_p^r are substantially larger than their counterparts \mathbf{W}_p^r . We point out that it is well-known that in the univariate case, typically, the behavior of the asymptotic characteristics of classes W_p^r and H_p^r coincide, even though H_p^r is a wider class than W_p^r . This phenomenon encouraged researchers to study the **B**-type classes $\mathbf{B}_{p,\theta}^r$ and determine the influence of the secondary parameter θ .

In Section 9 we provide some more arguments in favor of a systematic study of the hyperbolic cross approximation and classes of functions with mixed smoothness. In particular, we discuss the anisotropic mixed smoothness which plays an important role in high-dimensional approximation and applications; direct and inverse theorems for hyperbolic cross approximation; widths and hyperbolic cross approximation for the intersection of classes of mixed smoothness; continuous algorithms in m -term approximation and non-linear widths. We also comment on the quasi-Banach situation, i.e. the case $p < 1$. Corresponding classes of functions turned out to be suitable not only for best m -term approximation problems. In this section we also complement the results from Section 4 with recent results on other widths (s -numbers) like Weyl and Bernstein numbers relevant for the analysis of Monte Carlo algorithms. Also, we demonstrate there how classical hyperbolic cross approximation theory can be used in some important contemporary problems of numerical analysis.

High-dimensional approximation problems appear in several areas of science like for instance in quantum chemistry and meteorology. As already mentioned above some of our function class models are relevant in this context. In Section 10 we comment on some recent results on how the underlying dimension d affects the multivariate approximation error. The order of the approximation error is not longer sufficient for determining the *information based complexity* of

the problem. We present some recent results and techniques to see the d -dependence of the constants in the approximation error estimates and the convergence rate of widths complemented by sharp *preasymptotical* estimates in the Hilbert space case. In computational mathematics related to high-dimensional problems, the so-called ε -dimension $n_\varepsilon = n_\varepsilon(\mathbf{F}, X)$ is used to quantify the computational complexity. We discuss d -dependence of the ε -dimension of d -variate function classes of mixed smoothness and of the relevant hyperbolic cross approximations where d may be very large or even infinite, as well its application for numerical solving of parametric and stochastic elliptic PDEs.

The following list contains monographs and survey papers directly related to this paper: [203], [340], [316], [352], [69], [365], [45], [68], [253], [254], [255], [80], [248].

Notation. As usual \mathbb{N} denotes the natural numbers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{N}_{-1} := \mathbb{N}_0 \cup \{-1\}$, \mathbb{Z} denotes the integers, \mathbb{R} the real numbers, and \mathbb{C} the complex numbers. The letter d is always reserved for the underlying dimension in $\mathbb{R}^d, \mathbb{Z}^d$ etc. Elements $\mathbf{x}, \mathbf{y}, \mathbf{s} \in \mathbb{R}^d$ are always typesetted in bold face. We denote with (\mathbf{x}, \mathbf{y}) and $\mathbf{x} \cdot \mathbf{y}$ the usual Euclidean inner product in \mathbb{R}^d . For $a \in \mathbb{R}$ we denote $a_+ := \max\{a, 0\}$. For $0 < p \leq \infty$ and $\mathbf{x} \in \mathbb{R}^d$ we denote $|\mathbf{x}|_p := (\sum_{i=1}^d |x_i|^p)^{1/p}$ with the usual modification in the case $p = \infty$. By $\mathbf{x} = (x_1, \dots, x_d) > 0$ we mean that each coordinate is positive. By \mathbb{T} we denote the torus represented by the interval $[0, 2\pi]$. If X and Y are two (quasi-)normed spaces, the (quasi-)norm of an element x in X will be denoted by $\|x\|_X$. If $T : X \rightarrow Y$ is a continuous operator we write $T \in \mathcal{L}(X, Y)$. The symbol $X \hookrightarrow Y$ indicates that the identity operator is continuous. For two sequences a_n and b_n we will write $a_n \lesssim b_n$ if there exists a constant $c > 0$ such that $a_n \leq c b_n$ for all n . We will write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

Outstanding open problems

Let us conclude this introductory section with a list of outstanding open problems.

- 1.1 Find the right form of the Bernstein inequality for $\mathcal{T}(N)$ in L_1 (Open problem 2.1).
- 1.2 Prove the Small Ball Inequality for $d \geq 3$ (Open problem 2.5, 2.6).
- 1.3 Find the right order of the Kolmogorov widths $d_m(\mathbf{W}_p^r, L_\infty)$ and $d_m(\mathbf{H}_p^r, L_\infty)$ for $2 \leq p \leq \infty$ and $r > 1/p$ in dimension $d \geq 3$ (Open problem 4.2).
- 1.4 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{W}_p^r, L_p)$, $1 \leq p \leq \infty$, $r > 1/p$.
- 1.5 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{W}_p^r, L_q)$, $1 < p < 2 < q < \infty$.
- 1.6 Find the right order of the entropy numbers $\epsilon_n(\mathbf{W}_p^r, L_\infty)$ and $\epsilon_n(\mathbf{H}_p^r, L_\infty)$ for $1 \leq p \leq \infty$ and $r > 1/p$ in dimension $d \geq 3$ (Open problem 6.3).
- 1.7 Find the right order of the best m -term trigonometric approximation $\sigma_m(\mathbf{W}_p^r)_\infty$, $1 \leq p \leq \infty$ (Open problem 7.2).
- 1.8 Find the right order of the optimal error of numerical integration $\kappa_m(\mathbf{W}_\infty^r)$ (see Conjecture 8.29 in Subsection 8.9).
- 1.9 Find the right order of the optimal error of numerical integration $\kappa_m(\mathbf{W}_1^r)$ (see Conjecture 8.28 in Subsection 8.9).
- 1.10 Find the right order of the optimal error of numerical integration $\kappa_m(\mathbf{W}_p^r)$ in the range of small smoothness (see Conjectures 8.30, Subsection 8.9).

Acknowledgment. The authors acknowledge the fruitful discussions with D. Bazarkhanov, A. Hinrichs, T. Kühn, E. Novak and W. Sickel on this topic, especially at the ICERM Semester Programme “High-Dimensional Approximation” in Providence, 2014, where this project has been initiated, and at the conference “Approximation Methods and Function Spaces” in Hasenwinkel, 2015. The authors would further like to thank M. Hansen, J. Oettershagen, A.S. Romanyuk, S.A. Stasyuk, M. Ullrich and N. Temirgaliev for helpful comments on the material. V.N. Temlyakov and T. Ullrich would like to thank the organizers of the 2016 special semester “Constructive Approximation and Harmonic Analysis” at the Centre de Recerca Matemàtica (Barcelona) for the opportunity to present an advanced course based on this material. Moreover, T. Ullrich gratefully acknowledges support by the German Research Foundation (DFG), Ul-403/2-1, and the Emmy-Noether programme, Ul-403/1-1. D. Dũng thanks the Vietnam Institute for Advanced Study in Mathematics (VIASM) and the Vietnam National Foundation for Science and Technology Development (NAFOSTED), Grant No. 102.01-2014.02, for partial supports. D. Dũng and V.N. Temlyakov would like to thank T. Ullrich and M. Griebel for supporting their visits at the Institute for Numerical Simulation, University of Bonn, where major parts of this work were discussed. Also, D. Dũng and V.N. Temlyakov express their gratitude to the VIASM for support during Temlyakov’s visit of VIASM in September-October 2016, when certain parts of the survey were finished. Finally, the authors would like to thank G. Byrenheid, J. Oettershagen and S. Mayer (Bonn) for preparing most of the figures in the text.

2 Trigonometric polynomials

2.1 Univariate polynomials

Functions of the form

$$t(x) = \sum_{|k| \leq n} c_k e^{ikx} = a_0/2 + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx) \quad (2.1)$$

(c_k , a_k , b_k are complex numbers) will be called trigonometric polynomials of order n . The set of such polynomials we shall denote by $\mathcal{T}(n)$, and by $\mathcal{RT}(n)$ the subset of $\mathcal{T}(n)$ of real polynomials.

We first consider a number of concrete polynomials which play an important role in approximation theory.

1. The Dirichlet kernel. The Dirichlet kernel of order n :

$$\begin{aligned} \mathcal{D}_n(x) &:= \sum_{|k| \leq n} e^{ikx} = e^{-inx} (e^{i(2n+1)x} - 1) (e^{ix} - 1)^{-1} = \\ &= (\sin(n + 1/2)x) / \sin(x/2). \end{aligned}$$

The Dirichlet kernel is an even trigonometric polynomial with the majorant

$$|\mathcal{D}_n(x)| \leq \min\{2n + 1, \pi/|x|\}, \quad |x| \leq \pi. \quad (2.2)$$

The estimate

$$\|\mathcal{D}_n\|_1 \leq C \ln n, \quad n = 2, 3, \dots \quad (2.3)$$

follows from (2.2).

With $f * g$ we denote the convolution

$$f * g := (2\pi)^{-d} \int_{\mathbb{T}^d} f(\mathbf{y}) g(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

For any trigonometric polynomial $t \in \mathcal{T}(n)$ we have

$$t * \mathcal{D}_n = t,$$

Denote

$$x^l := 2\pi l / (2n + 1), \quad l = 0, 1, \dots, 2n.$$

Clearly, the points x^l , $l = 1, \dots, 2n$, are zeros of the Dirichlet kernel \mathcal{D}_n on $[0, 2\pi]$.

For any $t \in \mathcal{T}(n)$

$$t(x) = (2n + 1)^{-1} \sum_{l=0}^{2n} t(x^l) \mathcal{D}_n(x - x^l), \quad (2.4)$$

and for any $t \in \mathcal{T}(n)$,

$$\|t\|_2^2 = (2n + 1)^{-1} \sum_{l=0}^{2n} |t(x^l)|^2. \quad (2.5)$$

Sometimes it is convenient to consider the following slight modification of \mathcal{D}_n :

$$\mathcal{D}_n^1(x) := \mathcal{D}_n(x) - e^{-inx} = e^{-i(n-1)x} (e^{i2nx} - 1) (e^{ix} - 1)^{-1}. \quad (2.6)$$

Denote

$$y^l := \pi l/n, \quad l = 0, 1, \dots, 2n-1.$$

Clearly, the points y^l , $l = 1, \dots, 2n-1$, are zeros of the kernel \mathcal{D}_n^1 on $[0, 2\pi]$.

In the same way as (2.4) and (2.5) are proved for $t \in \mathcal{T}(n)$ one can prove the following identities for $t \in \mathcal{T}((-n, n]) := \text{span} \{e^{ikx}\}_{k=-n+1}^n$.

$$t(x) = (2n)^{-1} \sum_{l=0}^{2n-1} t(y^l) \mathcal{D}_n^1(x - y^l),$$

$$\|t\|_2^2 = (2n)^{-1} \sum_{l=0}^{2n-1} |t(y^l)|^2.$$

An advantage of \mathcal{D}_n^1 over \mathcal{D}_n is that the sets $\{y^l\}_{l=0}^{2^s-1}$, $s = 1, 2, \dots$, are nested.

The following relation for $1 < q < \infty$ is well known and easy to check

$$\|\mathcal{D}_n\|_q \asymp n^{1-1/q}. \quad (2.7)$$

The relation (2.7) for $q = \infty$ is obvious.

We denote by S_n the operator of taking the partial sum of order n . Then for $f \in L_1$ we have

$$S_n(f) = f * \mathcal{D}_n.$$

Theorem 2.1. *The operator S_n does not change polynomials from $\mathcal{T}(n)$ and for $p = 1$ or ∞ we have*

$$\|S_n\|_{p \rightarrow p} \leq C \ln n, \quad n = 2, 3, \dots,$$

and for $1 < p < \infty$ for all n we have

$$\|S_n\|_{p \rightarrow p} \leq C(p).$$

This theorem follows from (2.3) and the Marcinkiewicz multiplier theorem (see Theorem 11.10).

2. The Fejér kernel. The Fejér kernel of order $n-1$:

$$\begin{aligned} \mathcal{K}_{n-1}(x) &:= n^{-1} \sum_{m=0}^{n-1} \mathcal{D}_m(x) = \sum_{|m| \leq n} (1 - |m|/n) e^{imx} = \\ &= (\sin(nx/2))^2 / (n(\sin(x/2))^2). \end{aligned}$$

The Fejér kernel is an even nonnegative trigonometric polynomial in $\mathcal{T}(n-1)$ with the majorant

$$\mathcal{K}_{n-1}(x) \leq \min\{n, \pi^2/(nx^2)\}, \quad |x| \leq \pi.$$

From the obvious relations

$$\|\mathcal{K}_{n-1}\|_1 = 1, \quad \|\mathcal{K}_{n-1}\|_\infty = n$$

and the inequality

$$\|f\|_q \leq \|f\|_1^{1/q} \|f\|_\infty^{1-1/q}$$

we get

$$Cn^{1-1/q} \leq \|\mathcal{K}_{n-1}\|_q \leq n^{1-1/q}, \quad 1 \leq q \leq \infty.$$

3. The de la Vallée Poussin kernels. The de la Vallée Poussin kernels:

$$\mathcal{V}_{m,n}(x) := (n-m)^{-1} \sum_{l=m}^{n-1} \mathcal{D}_l(x), \quad n > m.$$

It is convenient to represent these kernels in terms of the Fejér kernels:

$$\begin{aligned} \mathcal{V}_{m,n}(x) &= (n-m)^{-1} (n\mathcal{K}_{n-1}(x) - m\mathcal{K}_{m-1}(x)) = \\ &= (\cos mx - \cos nx) (2(n-m)(\sin(x/2))^2)^{-1}. \end{aligned}$$

The de la Vallée Poussin kernels $\mathcal{V}_{m,n}$ are even trigonometric polynomials of order $n-1$ with the majorant

$$|\mathcal{V}_{m,n}(x)| \leq C \min\{n, 1/|x|, 1/((n-m)x^2)\}, \quad |x| \leq \pi. \quad (2.8)$$

The relation (2.8) implies the estimate

$$\|\mathcal{V}_{m,n}\|_1 \leq C \ln(1 + n/(n-m)).$$

We shall often use the de la Vallée Poussin kernel with $n = 2m$ and denote it by

$$\mathcal{V}_m(x) := \mathcal{V}_{m,2m}(x), \quad m \geq 1, \quad \mathcal{V}_0(x) = 1.$$

Then for $m \geq 1$ we have

$$\mathcal{V}_m = 2\mathcal{K}_{2m-1} - \mathcal{K}_{m-1},$$

which with the properties of \mathcal{K}_n implies

$$\|\mathcal{V}_m\|_1 \leq 3. \quad (2.9)$$

In addition

$$\|\mathcal{V}_m\|_\infty \leq 3m.$$

Consequently, in the same way as above we get

$$\|\mathcal{V}_m\|_q \asymp m^{1-1/q}, \quad 1 \leq q \leq \infty.$$

We denote

$$x(l) := \pi l / 2m, \quad l = 1, \dots, 4m.$$

Then as in (2.4) for each $t \in \mathcal{T}(m)$ we have

$$t(x) = (4m)^{-1} \sum_{l=1}^{4m} t(x(l)) \mathcal{V}_m(x - x(l)). \quad (2.10)$$

The operator V_m defined on L_1 by the formula

$$V_m(f) = f * \mathcal{V}_m$$

will be called the de la Vallée Poussin operator.

The following theorem is a corollary of the definition of kernels \mathcal{V}_m and the relation (2.9).

Theorem 2.2. *The operator V_m does not change polynomials from $\mathcal{T}(m)$ and for all $1 \leq p \leq \infty$ we have*

$$\|V_m\|_{p \rightarrow p} \leq 3, \quad m = 1, 2, \dots$$

4. The Rudin-Shapiro polynomials. For any natural number N there exists a polynomial of the form

$$\mathcal{R}_N(x) = \sum_{|k| \leq N} \varepsilon_k e^{ikx}, \quad \varepsilon_k = \pm 1,$$

such that the bound

$$\|\mathcal{R}_N\|_\infty \leq CN^{1/2}$$

holds.

2.2 Multivariate polynomials

The multivariate trigonometric system $\mathcal{T}^d := \mathcal{T} \times \cdots \times \mathcal{T}$, $d \geq 2$, in contrast to the univariate trigonometric system \mathcal{T} does not have a natural ordering. This leads to different natural ways of building sets of trigonometric polynomials. In this section we define the analogs of the Dirichlet, Fejér, de la Vallée Poussin and Rudin-Shapiro kernels for d -dimensional parallelepipeds

$$\Pi(\mathbf{N}, d) := \{\mathbf{a} \in \mathbb{R}^d : |a_j| \leq N_j, \ j = 1, \dots, d\},$$

where N_j are nonnegative integers. We shall formulate properties of these multivariate kernels, which easily follow from the corresponding properties of univariate kernels. Here $\mathcal{T}(\mathbf{N}, d)$ is the set of complex trigonometric polynomials with harmonics from $\Pi(\mathbf{N}, d)$. The set of real trigonometric polynomials with harmonics from $\Pi(\mathbf{N}, d)$ will be denoted by $\mathcal{RT}(\mathbf{N}, d)$. In the sequel we will frequently use the following notation

$$\nu(\bar{\mathbf{N}}) := \prod_{j=1}^d \bar{N}_j \quad \text{and} \quad \vartheta(\mathbf{N}) := \prod_{j=1}^d (2N_j + 1), \quad (2.11)$$

with $\bar{\mathbf{N}} := (\bar{N}_1, \dots, \bar{N}_d)$ and $\bar{N} := \max\{N, 1\}$.

1d. The Dirichlet kernels

$$\mathcal{D}_{\mathbf{N}}(\mathbf{x}) := \prod_{j=1}^d \mathcal{D}_{N_j}(x_j), \quad \mathbf{N} = (N_1, \dots, N_d)$$

have the following properties. For any trigonometric polynomial $t \in \mathcal{T}(\mathbf{N}, d)$,

$$t * \mathcal{D}_{\mathbf{N}} = t.$$

For $1 < q \leq \infty$,

$$\|\mathcal{D}_{\mathbf{N}}\|_q \asymp \nu(\bar{\mathbf{N}})^{1-1/q}$$

where $\bar{N}_j := \max\{N_j, 1\}$, $\nu(\bar{\mathbf{N}}) := \prod_{j=1}^d \bar{N}_j$ and

$$\|\mathcal{D}_{\mathbf{N}}\|_1 \asymp \prod_{j=1}^d \ln(N_j + 2).$$

We denote

$$P(\mathbf{N}) := \{\mathbf{n} = (n_1, \dots, n_d), \quad n_j \in \mathbb{N}_0, \quad 0 \leq n_j \leq 2N_j, \quad j = 1, \dots, d\},$$

and set

$$\mathbf{x}^{\mathbf{n}} := \left(\frac{2\pi n_1}{2N_1 + 1}, \dots, \frac{2\pi n_d}{2N_d + 1} \right), \quad \mathbf{n} \in P(\mathbf{N}).$$

Then for any $t \in \mathcal{T}(\mathbf{N}, d)$,

$$t(\mathbf{x}) = \vartheta(\mathbf{N})^{-1} \sum_{\mathbf{n} \in P(\mathbf{N})} t(\mathbf{x}^{\mathbf{n}}) \mathcal{D}_{\mathbf{N}}(\mathbf{x} - \mathbf{x}^{\mathbf{n}}),$$

where $\vartheta(\mathbf{N}) := \prod_{j=1}^d (2N_j + 1)$ and for any $t, u \in \mathcal{T}(\mathbf{N}, d)$,

$$\langle t, u \rangle = \vartheta(\mathbf{N})^{-1} \sum_{\mathbf{n} \in P(\mathbf{N})} t(\mathbf{x}^{\mathbf{n}}) \bar{u}(\mathbf{x}^{\mathbf{n}}),$$

$$\|t\|_2^2 = \vartheta(\mathbf{N})^{-1} \sum_{\mathbf{n} \in P(\mathbf{N})} |t(\mathbf{x}^{\mathbf{n}})|^2.$$

2d. The Fejér kernels

$$\mathcal{K}_{\mathbf{N}}(\mathbf{x}) := \prod_{j=1}^d \mathcal{K}_{N_j}(x_j), \quad \mathbf{N} = (N_1, \dots, N_d),$$

are nonnegative trigonometric polynomials from $\mathcal{T}(\mathbf{N}, d)$, which have the following properties (recall (2.11)):

$$\begin{aligned} \|\mathcal{K}_{\mathbf{N}}\|_1 &= 1, \\ \|\mathcal{K}_{\mathbf{N}}\|_q &\asymp \vartheta(\mathbf{N})^{1-1/q}, \quad 1 \leq q \leq \infty, \\ \|\mathcal{K}_{\mathbf{N}}\|_{\mathbf{q}} &\asymp \prod_{j=1}^d (\max\{1, N_j\})^{1-1/q_j}, \quad \mathbf{1} \leq \mathbf{q} \leq \infty, \mathbf{q} = (q_1, \dots, q_d). \end{aligned}$$

3d. The de la Vallée Poussin kernels

$$\mathcal{V}_{\mathbf{N}}(\mathbf{x}) := \prod_{j=1}^d \mathcal{V}_{N_j}(x_j), \quad \mathbf{N} = (N_1, \dots, N_d),$$

have the following properties (recall (2.11))

$$\begin{aligned} \|\mathcal{V}_{\mathbf{N}}\|_1 &\leq 3^d, \\ \|\mathcal{V}_{\mathbf{N}}\|_q &\asymp \vartheta(\mathbf{N})^{1-1/q}, \quad 1 \leq q \leq \infty, \\ \|\mathcal{V}_{\mathbf{N}}\|_{\mathbf{q}} &\asymp \prod_{j=1}^d (\max\{1, N_j\})^{1-1/q_j}, \quad \mathbf{1} \leq \mathbf{q} \leq \infty. \end{aligned} \tag{2.12}$$

For any $t \in \mathcal{T}(\mathbf{N}, d)$,

$$V_{\mathbf{N}}(t) := t * \mathcal{V}_{\mathbf{N}} = t.$$

We denote

$$P'(\mathbf{N}) := \{\mathbf{n} = (n_1, \dots, n_d), \quad n_j \in \mathbb{N}_0, \quad 1 \leq n_j \leq 4N_j, \quad j = 1, \dots, d\}$$

and set

$$\mathbf{x}(\mathbf{n}) := \left(\frac{\pi n_1}{2N_1}, \dots, \frac{\pi n_d}{2N_d} \right), \quad \mathbf{n} \in P'(\mathbf{N}).$$

In the case $N_j = 0$ we assume $x_j(\mathbf{n}) = 0$. Then for any $t \in \mathcal{T}(\mathbf{N}, d)$ we have the representation (recall (2.11))

$$t(\mathbf{x}) = \nu(4\bar{\mathbf{N}})^{-1} \sum_{\mathbf{n} \in P'(\mathbf{N})} t(\mathbf{x}(\mathbf{n})) \mathcal{V}_{\mathbf{N}}(\mathbf{x} - \mathbf{x}(\mathbf{n})). \quad (2.13)$$

The relation (2.12) implies that

$$\|V_{\mathbf{N}}\|_{\mathbf{p} \rightarrow \mathbf{p}} \leq 3^d, \quad \mathbf{1} \leq \mathbf{p} \leq \infty.$$

Let us define the polynomials $\mathcal{A}_{\mathbf{s}}(\mathbf{x})$ for $\mathbf{s} = (s_1, \dots, s_d) \in \mathbb{N}_0^d$

$$\mathcal{A}_{\mathbf{s}}(\mathbf{x}) := \prod_{j=1}^d \mathcal{A}_{s_j}(x_j),$$

with $\mathcal{A}_{s_j}(x_j)$ defined as follows:

$$\mathcal{A}_0(x) := 1, \quad \mathcal{A}_1(x) := \mathcal{V}_1(x) - 1, \quad \mathcal{A}_s(x) := \mathcal{V}_{2^{s-1}}(x) - \mathcal{V}_{2^{s-2}}(x), \quad s \geq 2,$$

where \mathcal{V}_m are the de la Vallée Poussin kernels. Then by (2.9),

$$\|\mathcal{A}_{\mathbf{s}}(\mathbf{x})\|_1 \leq 6^d,$$

and consequently we have for the operator $A_{\mathbf{s}}$, which is the convolution with the kernel $\mathcal{A}_{\mathbf{s}}(\mathbf{x})$, the inequality

$$\|A_{\mathbf{s}}\|_{\mathbf{p} \rightarrow \mathbf{p}} \leq 6^d, \quad \mathbf{1} \leq \mathbf{p} \leq \infty. \quad (2.14)$$

We note that in the case $\mathbf{s} \geq \mathbf{2}$ for any $t \in \mathcal{T}(2^{\mathbf{s}-\mathbf{2}}, d)$,

$$A_{\mathbf{s}}(t) = 0.$$

4d. The Rudin–Shapiro polynomials

$$\mathcal{R}_{\mathbf{N}}(\mathbf{x}) := \prod_{j=1}^d \mathcal{R}_{N_j}(x_j), \quad \mathbf{N} = (N_1, \dots, N_d),$$

have the following properties: $\mathcal{R}_{\mathbf{N}} \in \mathcal{T}(\mathbf{N}, d)$,

$$\|\mathcal{R}_{\mathbf{N}}\|_{\infty} \leq C(d) \vartheta(\mathbf{N})^{1/2}, \quad \hat{\mathcal{R}}_{\mathbf{N}}(\mathbf{k}) = \pm 1, \quad |\mathbf{k}| \leq \mathbf{N}.$$

The Rudin-Shapiro polynomials have all the Fourier coefficients with their absolute values equal to one. This is similar to the Dirichlet kernels. However, the L_p norms of $\mathcal{R}_{\mathbf{N}}$ behave in a very different way:

$$\|\mathcal{R}_{\mathbf{N}}\|_p \asymp \vartheta(\mathbf{N})^{1/2}, \quad 1 \leq p \leq \infty.$$

In some applications we need to construct polynomials with similar properties in a subspace of the $\mathcal{T}(\mathbf{N}, d)$. We present here one known result in that direction (see [352], Ch.2, Theorem 1.1 and [344]).

Theorem 2.3. *Let $\varepsilon > 0$ and a subspace $\Psi \subset \mathcal{T}(\mathbf{N}, d)$ be such that $\dim \Psi \geq \varepsilon \vartheta(\mathbf{N})$. Then there is a $t \in \Psi$ such that*

$$\|t\|_{\infty} = 1, \quad \|t\|_2 \geq C(\varepsilon, d) > 0.$$

2.3 Hyperbolic cross polynomials

Let $\mathbf{s} = (s_1, \dots, s_d)$ be a vector whose coordinates are nonnegative integers

$$\begin{aligned} \rho(\mathbf{s}) &:= \{\mathbf{k} \in \mathbb{Z}^d : [2^{s_j-1}] \leq |k_j| < 2^{s_j}, \quad j = 1, \dots, d\}, \\ Q_n &:= \cup_{|\mathbf{s}|_1 \leq n} \rho(\mathbf{s}) \quad - \text{ a step hyperbolic cross}, \\ \Gamma(N) &:= \{\mathbf{k} \in \mathbb{Z}^d : \prod_{j=1}^d \max\{|k_j|, 1\} \leq N\} \quad - \text{ a hyperbolic cross}. \end{aligned} \tag{2.15}$$

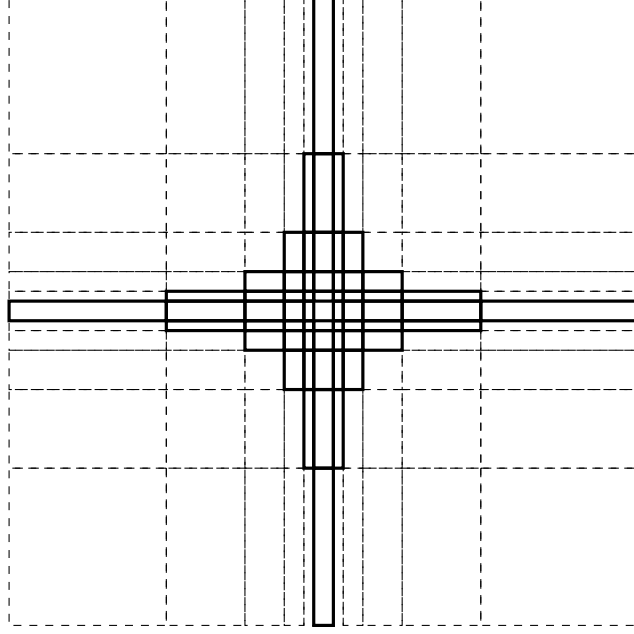


Figure 2: A step hyperbolic cross Q_n in $d = 2$

For $f \in L_1(\mathbb{T}^d)$

$$\delta_{\mathbf{s}}(f, \mathbf{x}) := \sum_{\mathbf{k} \in \rho(\mathbf{s})} \hat{f}(\mathbf{k}) e^{i(\mathbf{k}, \mathbf{x})}.$$

Let G be a finite set of points in \mathbb{Z}^d , we denote

$$\mathcal{T}(G) := \left\{ t : t(\mathbf{x}) = \sum_{\mathbf{k} \in G} c_{\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \right\}.$$

For the sake of simplicity we shall write $\mathcal{T}(\Gamma(N)) = \mathcal{T}(N)$. The unit L_p -ball in $\mathcal{T}(G)$ we denote by $\mathcal{T}(G)_p$ and in addition

$$\mathcal{T}^\perp(G) := \{g \in L_1 : \text{ for all } f \in \mathcal{T}(G), \quad \langle f, g \rangle = 0\}.$$

As above for $G = \Gamma(N)$ we write $\mathcal{T}^\perp(N)$ instead of $\mathcal{T}^\perp(\Gamma(N))$. We shall use the following

simple relations (recall the notation in (2.11))

$$\begin{aligned}
|\Gamma(N)| &\asymp N(\log N)^{d-1}; & |Q_n| &\asymp 2^n n^{d-1}; \\
\sum_{\mathbf{k} > \mathbf{0}, \nu(\mathbf{k}) > N} \nu(\mathbf{k})^{-r} &\asymp N^{-r+1}(\log N)^{d-1}, & r > 1; \\
\sum_{|\mathbf{s}|_1 > n} 2^{-r|\mathbf{s}|_1} &\asymp 2^{-rn} n^{d-1}, & r > 0.
\end{aligned} \tag{2.16}$$

Note, that the sum in the middle can be rewritten (via dyadic blocks) to a sum $\sum_{|\mathbf{s}|_1 > n} 2^{-(r-1)|\mathbf{s}|_1}$ with $|n - \log N| \leq c$. Sums of this type have been discussed in detail in Lemmas A – D in the introduction of [340]. Refined estimates for the cardinality of hyperbolic crosses of any kind in high dimensions can be found in the recent papers [210, Lem. 3.1, 3.2, Thm. 4.9] and [56], see also (10.3) in Section 10 below.

It is easy to see that

$$Q_n \subset \Gamma(2^n) \subset Q_{n+d}.$$

Therefore it is enough to prove a number of properties of polynomials such as the Bernstein and Nikol'skii inequalities for one set $\mathcal{T}(Q_n)$ or $\mathcal{T}(N)$ only.

We shall consider the following trigonometric polynomials.

1h. The analogs of the Dirichlet kernels. Consider

$$\mathcal{D}_{Q_n}(\mathbf{x}) := \sum_{\mathbf{k} \in Q_n} e^{i(\mathbf{k}, \mathbf{x})} = \sum_{|\mathbf{s}|_1 \leq n} \mathcal{D}_{\rho(\mathbf{s})}(\mathbf{x}),$$

where $\mathcal{D}_{\rho(\mathbf{s})}(\mathbf{x}) := \sum_{\mathbf{k} \in \rho(\mathbf{s})} e^{i(\mathbf{k}, \mathbf{x})}$. It is clear that for $t \in \mathcal{T}(Q_n)$,

$$t * \mathcal{D}_{Q_n} = t.$$

We have the following behavior of the L_p norms of the Dirichlet kernels (see [352], Ch.3, Lemma 1.1).

Lemma 2.4. *Let $1 < p < \infty$. Then*

$$\|\mathcal{D}_{Q_n}(\mathbf{x})\|_p \asymp 2^{(1-1/p)n} n^{(d-1)/p}.$$

2h. The analogs of the de la Vallée Poussin kernels. Let $\mathcal{A}_{\mathbf{s}}(x)$ be the polynomials which have been defined above. These polynomials are from $\mathcal{T}(2^{\mathbf{s}}, d)$ and

$$\hat{\mathcal{A}}_{\mathbf{s}}(\mathbf{k}) \neq 0 \quad \text{only for} \quad \mathbf{k} : \quad 2^{\mathbf{s}_j-2} < |k_j| < 2^{\mathbf{s}_j}, \quad j = 1, \dots, d.$$

We define the polynomials

$$\mathcal{V}_{Q_n}(\mathbf{x}) := \sum_{|\mathbf{s}|_1 \leq n} \mathcal{A}_{\mathbf{s}}(\mathbf{x}).$$

These are polynomials in $\mathcal{T}(Q_n)$ with the property

$$\hat{\mathcal{V}}_{Q_{n+d}}(\mathbf{k}) = 1 \quad \text{for} \quad \mathbf{k} \in Q_n.$$

We shall use the following notation. Let $f \in L_1$

$$S_{Q_n}(f) := f * \mathcal{D}_{Q_n},$$

$$V_{Q_n}(f) := f * \mathcal{V}_{Q_n},$$

$$A_s(f) := f * \mathcal{A}_s.$$

From Corollary 11.8 to the Littlewood-Paley theorem (see Appendix) it follows that for $1 < p < \infty$

$$\|S_{Q_n}\|_{p \rightarrow p} \leq C(p, d).$$

In Subsection 2.2 it was established that the L_1 -norms of the de la Vallée Poussin kernels for parallelepipeds are uniformly bounded. This fact plays an essential role in studying approximation problems in the L_1 and L_∞ norms. The following lemma shows that, unfortunately, the kernels \mathcal{V}_{Q_n} have no such property (see [352], Ch.3, Lemma 1.2).

Lemma 2.5. *Let $1 \leq p < \infty$. Then the following relation*

$$\|\mathcal{V}_{Q_n}(\mathbf{x})\|_p \asymp 2^{(1-1/p)n} n^{(d-1)/p}.$$

holds.

Lemma 2.5 highlights a new phenomenon for hyperbolic cross polynomials – there is no analogs of the de la Vallée Poussin kernels for the hyperbolic crosses with uniformly bounded L_1 norms. In particular, it follows from the inequality: For any $\epsilon > 0$ there is a number C_ϵ such that for all $t \in \mathcal{T}(N)$ (see [340], Ch.1, Section 2)

$$\sum_{\mathbf{k} \in \Gamma(N)} |\hat{t}(\mathbf{k})| \leq C_\epsilon (\ln N)^\epsilon N \|t\|_1. \quad (2.17)$$

This new phenomenon substantially complicates the study of approximation by hyperbolic cross polynomials in the L_1 and L_∞ norms. The reader can find a discussion of related questions in [340], Chapter 2, Section 5.

2.4 The Bernstein-Nikol'skii inequalities

1. The Bernstein inequalities. We define the operator D_α^r , $r \geq 0$, $\alpha \in \mathbb{R}$, on the set of trigonometric polynomials as follows: let $f \in \mathcal{T}(n)$; then

$$D_\alpha^r f = f^{(r)}(x, \alpha) := f(x) * \mathcal{V}_n^r(x, \alpha), \quad (2.18)$$

$$\mathcal{V}_n^r(x, \alpha) := 1 + 2 \sum_{k=1}^n k^r \cos(kx + \alpha\pi/2) + 2 \sum_{k=n+1}^{2n-1} k^r (1 - (k-n)/n) \cos(kx + \alpha\pi/2) \quad (2.19)$$

and $f^{(r)}(x, \alpha)$ will be called the (r, α) derivative. It is clear that for $f(x)$ such that $\hat{f}(0) = 0$ we have for natural numbers r ,

$$D_r^r f = \frac{d^r}{dx^r} f.$$

The operator D_α^r is defined in such a way that it has an inverse operator for each $\mathcal{T}(n)$. This property distinguishes D_α^r from the differential operator and it will be convenient for us. On the other hand it is clear that

$$\frac{d^r f}{dx^r} = D_r^r f - \hat{f}(0).$$

Theorem 2.6. *For any $t \in \mathcal{T}(n)$ we have ($r > 0$, $\alpha \in \mathbb{R}$, $1 \leq p \leq \infty$)*

$$\|t^{(r)}(x, \alpha)\|_p \leq C(r) n^r \|t\|_p, \quad n = 1, 2, \dots$$

Theorem 2.6 can be easily generalized to the multidimensional case of trigonometric polynomials from $\mathcal{T}(\mathbf{N}, d)$. Let $\mathbf{r} = (r_1, \dots, r_d)$, $r_j \geq 0$, $j = 1, \dots, d$, $\alpha = (\alpha_1, \dots, \alpha_d)$, $\mathbf{N} = (N_1, \dots, N_d)$. We consider the polynomials

$$\mathcal{V}_{\mathbf{N}}^{\mathbf{r}}(\mathbf{x}, \alpha) = \prod_{j=1}^d \mathcal{V}_{N_j}^{r_j}(x_j, \alpha_j),$$

where $\mathcal{V}_{N_j}^{r_j}(x_j, \alpha_j)$ are defined in (2.19).

We define the operator $D_{\alpha}^{\mathbf{r}}$ on the set of trigonometric polynomials as follows: let $f \in \mathcal{T}(\mathbf{N}, d)$, then

$$D_{\alpha}^{\mathbf{r}} f := f^{(\mathbf{r})}(\mathbf{x}, \alpha) := f(\mathbf{x}) * \mathcal{V}_{\mathbf{N}}^{\mathbf{r}}(\mathbf{x}, \alpha),$$

and we shall call $D_{\alpha}^{\mathbf{r}} f$ the (\mathbf{r}, α) -derivative. In the case of identical components $r_j = r$, $j = 1, \dots, d$, we shall write the scalar r in place of the vector.

Theorem 2.7. *Let $\mathbf{r} \geq \mathbf{0}$ and $\alpha \in \mathbb{R}^d$ be such that for $r_j = 0$ we have $\alpha_j = 0$. Then for any $t \in \mathcal{T}(\mathbf{N}, d)$, $\mathbf{N} > \mathbf{0}$, the inequality*

$$\|t^{(\mathbf{r})}(\cdot, \alpha)\|_{\mathbf{p}} \leq C(\mathbf{r}) \|t\|_{\mathbf{p}} \prod_{j=1}^d N_j^{r_j}, \quad \mathbf{1} \leq \mathbf{p} \leq \infty,$$

holds.

It is easy to check that the above Bernstein inequalities are sharp. Extension of Theorem 2.7 to the case of hyperbolic cross polynomials is nontrivial and brings out a new phenomenon.

Theorem 2.8. *For arbitrary α*

$$\sup_{t \in \mathcal{T}(\mathbf{N})} \|t^{(\mathbf{r})}(\mathbf{x}, \alpha)\|_p / \|t\|_p \asymp \begin{cases} N^r & \text{for } 1 < p < \infty, \quad r \geq 0. \\ N^r (\log N)^{d-1} & \text{for } p = \infty, \quad r > 0. \end{cases}$$

The Bernstein inequalities in Theorem 2.8 have different form for $p = \infty$ and $1 < p < \infty$. The upper bound in the case $p = \infty$ was obtained by Babenko [7]. The matching lower bound for $p = \infty$ was proved by Telyakovskii [314]. The case $1 < p < \infty$ was settled by Mityagin [233]. The right form of the Bernstein inequalities in case $p = 1$ is not known. It was proved in [357] that in the case $d = 2$

$$\sup_{t \in \mathcal{T}(\mathbf{N})} \|t^{(\mathbf{r})}(\mathbf{x}, \alpha)\|_1 / \|t\|_1 \lesssim (\ln N)^{1/2} N^r.$$

2. The Nikol'skii inequalities. The following inequalities are well known and easy to prove.

Theorem 2.9. *For any $t \in \mathcal{T}(n)$, $n > 0$, we have the inequality*

$$\|t\|_p \leq C n^{1/q-1/p} \|t\|_q, \quad 1 \leq q < p \leq \infty.$$

The above univariate inequalities can be extended to the case of polynomials from $\mathcal{T}(\mathbf{N}, d)$.

Theorem 2.10. *For any $t \in \mathcal{T}(\mathbf{N}, d)$, $\mathbf{N} > \mathbf{0}$ the following inequality holds ($\mathbf{1} \leq \mathbf{q} \leq \mathbf{p} \leq \infty$):*

$$\|t\|_{\mathbf{p}} \leq C(d) \|t\|_{\mathbf{q}} \prod_{j=1}^d N_j^{1/q_j-1/p_j}.$$

We formulate the above inequalities for vector \mathbf{p} and \mathbf{q} because in this form they are used to prove embedding type inequalities. We proceed to the problem of estimating $\|f\|_p$ in terms of the array $\{\|\delta_{\mathbf{s}}(f)\|_q\}_{\mathbf{s} \in \mathbb{N}_0^d}$ where Theorem 2.10 is heavily used. Here and below p and q are scalars such that $1 \leq q, p \leq \infty$. Let an array $\varepsilon = \{\varepsilon_{\mathbf{s}}\}$ be given, where $\varepsilon_{\mathbf{s}} \geq 0$, $\mathbf{s} = (s_1, \dots, s_d)$, and s_j are nonnegative integers, $j = 1, \dots, d$. We denote by $G(\varepsilon, q)$ and $F(\varepsilon, q)$ the following sets of functions ($1 \leq q \leq \infty$):

$$G(\varepsilon, q) := \{f \in L_q : \|\delta_{\mathbf{s}}(f)\|_q \leq \varepsilon_{\mathbf{s}} \quad \text{for all } \mathbf{s}\},$$

$$F(\varepsilon, q) := \{f \in L_q : \|\delta_{\mathbf{s}}(f)\|_q \geq \varepsilon_{\mathbf{s}} \quad \text{for all } \mathbf{s}\}.$$

Theorem 2.11. *The following relations hold:*

$$\sup_{f \in G(\varepsilon, q)} \|f\|_p \asymp \left(\sum_{\mathbf{s}} \varepsilon_{\mathbf{s}}^p 2^{|\mathbf{s}|_1(p/q-1)} \right)^{1/p}, \quad 1 \leq q < p < \infty; \quad (2.20)$$

$$\inf_{f \in F(\varepsilon, q)} \|f\|_p \asymp \left(\sum_{\mathbf{s}} \varepsilon_{\mathbf{s}}^p 2^{|\mathbf{s}|_1(p/q-1)} \right)^{1/p}, \quad 1 < p < q \leq \infty, \quad (2.21)$$

with constants independent of ε .

Theorem 2.11 was proved in [334] (see also [340], Ch.1, Theorem 3.3). Theorem 2.11 can be formulated in the form of embeddings: relation (2.20) implies Lemma 3.13 and relation (2.21) implies Lemma 3.14 (see Section 3 below).

Remark 2.12. *In the proof of the upper bound in (2.20) from [340] we used only the property $\delta_{\mathbf{s}}(f) \in \mathcal{T}(2^{\mathbf{s}}, d)$. That is, if*

$$f = \sum_{\mathbf{s}} t_{\mathbf{s}}, \quad t_{\mathbf{s}} \in \mathcal{T}(2^{\mathbf{s}}, d),$$

then for $1 \leq q < p < \infty$,

$$\|f\|_p \leq C(q, p, d) \left(\sum_{\mathbf{s}} \|t_{\mathbf{s}}\|_q^p 2^{|\mathbf{s}|_1(p/q-1)} \right)^{1/p}. \quad (2.22)$$

The above Remark 2.12 is from [91]. This remark is very useful in studying sampling recovery by Smolyak's algorithms (see Section 5).

The Nikol'skii inequalities for polynomials from $\mathcal{T}(N)$ are nontrivial in the case $q = 1$. The following two theorems are from [340], Ch.1, Section 2.

Theorem 2.13. *Suppose that $1 \leq q < \infty$ and $r \geq 0$. Then*

$$\sup_{t \in \mathcal{T}(N)} \|t^{(r)}(\mathbf{x}, \alpha)\|_{\infty} / \|t\|_q \asymp N^{r+1/q} (\log N)^{(d-1)(1-1/q)}.$$

Theorem 2.14. *Suppose that $1 \leq q \leq p < \infty$, $p > 1$, $r \geq 0$. Then*

$$\sup_{t \in \mathcal{T}(N)} \|t^{(r)}(\mathbf{x}, \alpha)\|_p / \|t\|_q \asymp N^{r+1/q-1/p}.$$

3. The Marcinkiewicz theorem. The set $\mathcal{T}(n)$ of trigonometric polynomials is a space of dimension $2n + 1$. Each polynomial $t \in \mathcal{T}(n)$ is uniquely defined by its Fourier coefficients $\{\hat{t}(k)\}_{|k| \leq n}$ and by the Parseval identity we have

$$\|t\|_2^2 = \sum_{|k| \leq n} |\hat{t}(k)|^2,$$

which means that the set $\mathcal{T}(n)$ as a subspace of L_2 is isomorphic to ℓ_2^{2n+1} . The relation (2.5) shows that a similar isomorphism can be set up in another way: mapping a polynomial $t \in \mathcal{T}(n)$ to the vector $m(t) := \{t(x^l)\}_{l=0}^{2n}$ of its values at the points

$$x^l = 2\pi l / (2n + 1), \quad l = 0, \dots, 2n.$$

The relation (2.5) gives

$$\|t\|_2 = (2n + 1)^{-1/2} \|m(t)\|_2.$$

The following statement is the Marcinkiewicz theorem.

Theorem 2.15. *Let $1 < p < \infty$; then for $t \in \mathcal{T}(n)$, $n > 0$, we have the relation*

$$C_1(p) \|t\|_p \leq n^{-1/p} \|m(t)\|_p \leq C_2(p) \|t\|_p.$$

The following statement is analogous to Theorem 2.15 but in contrast to it includes the cases $p = 1$ and $p = \infty$.

Theorem 2.16. *Let $x(l) = \pi l / (2n)$, $l = 1, \dots, 4n$, $M(t) := (t(x(1)), \dots, t(x(4n)))$. Then for an arbitrary $t \in \mathcal{T}(n)$, $n > 0$, $1 \leq p \leq \infty$,*

$$C_1 \|t\|_p \leq n^{-1/p} \|M(t)\|_p \leq C_2 \|t\|_p.$$

Similar inequalities hold for polynomials from $\mathcal{T}(\mathbf{N}, d)$. We formulate the equivalence of a mixed norm of a trigonometric polynomial to its mixed lattice norm. We use the notation

$$\ell(\mathbf{N}, d) = \{\mathbf{a} = \{a_{\mathbf{n}}\}, \quad \mathbf{n} = (n_1, \dots, n_d), \quad 1 \leq n_j \leq N_j, \quad j = 1, \dots, d\},$$

and for $\mathbf{a} \in \ell(\mathbf{N}, d)$ we define the mixed norm

$$\begin{aligned} \|\mathbf{a}\|_{\mathbf{p}, \mathbf{N}} &:= \left(\sum_{n_d=1}^{N_d} N_d^{-1} \left(\dots \left(\sum_{n_1=1}^{N_1} N_1^{-1} |a_{\mathbf{n}}|^{p_1} \right)^{p_2/p_1} \dots \right)^{p_d/p_{d-1}} \right)^{1/p_d} \\ &= \|\mathbf{a}\|_{\mathbf{p}} \prod_{j=1}^d N_j^{-1/p_j}. \end{aligned}$$

Theorem 2.17. *Let $t \in \mathcal{T}(\mathbf{N}, d)$, $\mathbf{N} > \mathbf{0}$. Then for any $\mathbf{1} \leq \mathbf{p} \leq \infty$,*

$$C_d^{-1} \left\| \{t(\mathbf{x}(\mathbf{n}))\}_{\mathbf{n} \in P'(\mathbf{N})} \right\|_{\mathbf{p}, 4\mathbf{N}} \leq \|t\|_{\mathbf{p}} \leq C_d \left\| \{t(\mathbf{x}(\mathbf{n}))\}_{\mathbf{n} \in P'(\mathbf{N})} \right\|_{\mathbf{p}, 4\mathbf{N}},$$

where C_d is a number depending only on d .

In the case $\mathbf{p} = (p, \dots, p)$ this theorem is an immediate corollary of the corresponding one-dimensional Theorem 2.16.

It turns out that there is no analog of Theorem 2.17 for polynomials from $\mathcal{T}(N)$. We discuss this issue in the next section.

2.5 Volume estimates

Let $\mathbf{s} = (s_1, \dots, s_d)$ be a vector with nonnegative integer coordinates ($\mathbf{s} \in \mathbb{N}_0^d$). Denote for a natural number n

$$\Delta Q_n := Q_n \setminus Q_{n-1} = \cup_{|\mathbf{s}|_1=n} \rho(\mathbf{s})$$

with $|\mathbf{s}|_1 = s_1 + \dots + s_d$ for $\mathbf{s} \in \mathbb{N}_0^d$. We call a set ΔQ_n *hyperbolic layer*. For a set $\Lambda \subset \mathbb{Z}^d$ denote

$$\mathcal{T}(\Lambda) := \{f \in L_1 : \hat{f}(\mathbf{k}) = 0, \mathbf{k} \in \mathbb{Z}^d \setminus \Lambda\}.$$

For a finite set Λ we assign to each $f = \sum_{\mathbf{k} \in \Lambda} \hat{f}(\mathbf{k}) e^{i(\mathbf{k}, \mathbf{x})} \in \mathcal{T}(\Lambda)$ a vector

$$A(f) := \{(\operatorname{Re} \hat{f}(\mathbf{k}), \operatorname{Im} \hat{f}(\mathbf{k})), \quad \mathbf{k} \in \Lambda\} \in \mathbb{R}^{2|\Lambda|}$$

where $|\Lambda|$ denotes the cardinality of Λ and define

$$B_\Lambda(L_p) := \{A(f) : f \in \mathcal{T}(\Lambda), \quad \|f\|_p \leq 1\}.$$

In the case $\Lambda = \Pi(\mathbf{N}) := \Pi(\mathbf{N}, d) := [-N_1, N_1] \times \dots \times [-N_d, N_d]$, $\mathbf{N} := (N_1, \dots, N_d)$, the following volume estimates are known.

Theorem 2.18. *For any $1 \leq p \leq \infty$ we have*

$$(\operatorname{vol}(B_{\Pi(\mathbf{N})}(L_p)))^{(2|\Pi(\mathbf{N})|)^{-1}} \asymp |\Pi(\mathbf{N})|^{-1/2},$$

with constants in \asymp that may depend only on d .

We note that the most difficult part of Theorem 2.18 is the lower estimate for $p = \infty$. The corresponding estimate was proved in the case $d = 1$ in [193] and in the general case in [344] and [350]. The upper estimate for $p = 1$ in Theorem 2.18 can be easily reduced to the volume estimate for an octahedron (see, for instance [348]).

The results of [195] imply the following estimate.

Theorem 2.19. *For any finite set $\Lambda \subset \mathbb{Z}^d$ and any $1 \leq p \leq 2$ we have*

$$\operatorname{vol}(B_\Lambda(L_p))^{(2|\Lambda|)^{-1}} \asymp |\Lambda|^{-1/2}.$$

The following result of Bourgain-Milman [41] plays an important role in the volume estimates of finite dimensional bodies.

Theorem 2.20. *For any convex centrally symmetric body $K \subset \mathbb{R}^n$ we have*

$$(\operatorname{vol}(K) \operatorname{vol}(K^o))^{1/n} \asymp (\operatorname{vol}(B_2^n))^{2/n} \asymp 1/n$$

where K^o is a polar for K , that is

$$K^o := \{\mathbf{x} \in \mathbb{R}^n : \sup_{\mathbf{y} \in K} (\mathbf{x}, \mathbf{y}) \leq 1\}.$$

Remark 2.21. *For the case of ℓ_p^n balls with $1 \leq p \leq \infty$ or, more general, unit balls of symmetric norms on \mathbb{R}^n and their duals, Theorem 2.20 has been proved earlier by Schütt [285].*

The following result is from [199].

Theorem 2.22. *Let Λ have the form $\Lambda = \cup_{\mathbf{s} \in S} \rho(\mathbf{s})$, $S \subset \mathbb{N}_0^d$ is a finite set. Then for any $1 \leq p < \infty$ we have*

$$\operatorname{vol}(B_\Lambda(L_p))^{(2|\Lambda|)^{-1}} \asymp |\Lambda|^{-1/2}.$$

We now proceed to results for the case $d = 2$. Denote $N := 2|\Delta Q_n|$. Let

$$E_\Lambda^\perp(f)_p := \inf_{g \perp \mathcal{T}(\Lambda)} \|f - g\|_p,$$

$$B_\Lambda^\perp(L_p) := \{A(f) : f \in \mathcal{T}(\Lambda), \quad E_\Lambda^\perp(f)_p \leq 1\}.$$

Theorem 2.23. *In the case $d = 2$ we have for $N := 2|\Delta Q_n|$*

$$(\text{vol}(B_{\Delta Q_n}(L_\infty)))^{1/N} \asymp (2^n n^2)^{-1/2}; \quad (2.23)$$

$$(\text{vol}(B_{\Delta Q_n}^\perp(L_1)))^{1/N} \asymp 2^{-n/2}.$$

It is interesting to compare the first relation in Theorem 2.23 with the following estimate for $1 \leq p < \infty$ that follows from Theorem 2.22

$$(\text{vol}(B_{\Delta Q_n}(L_p)))^{1/N} \asymp (2^n n)^{-1/2}.$$

We see that in the case $\Lambda = \Delta Q_n$ unlike the case $\Lambda = \Pi(N_1, \dots, N_d)$ the estimate for $p = \infty$ is different from the estimate for $1 \leq p < \infty$.

The discrete L_∞ -norm for polynomials from $\mathcal{T}(\Lambda)$.

We present here some results from [199] (see also [197] and [198]). We begin with the following conditional statement.

Theorem 2.24. *Assume that a finite set $\Lambda \subset \mathbb{Z}^d$ has the following properties:*

$$(\text{vol}(B_\Lambda(L_\infty)))^{1/N} \leq K_1 N^{-1/2}, \quad N := 2|\Lambda|,$$

and a set $\Omega = \{x^1, \dots, x^M\}$ satisfies the condition

$$\forall f \in \mathcal{T}(\Lambda) \quad \|f\|_\infty \leq K_2 \|f\|_{\infty, \Omega}, \quad \|f\|_{\infty, \Omega} := \max_{x \in \Omega} |f(x)|.$$

Then there exists an absolute constant $C > 0$ such that

$$M \geq N e^{C(K_1 K_2)^{-2}}.$$

We now give some corollaries from Theorem 2.24.

Theorem 2.25. *Assume a finite set $\Omega \subset \mathbb{T}^2$ has the following property.*

$$\forall t \in \mathcal{T}(\Delta Q_n) \quad \|t\|_\infty \leq K_2 \|t\|_{\infty, \Omega}. \quad (2.24)$$

Then

$$|\Omega| \geq 2|\Delta Q_n| e^{Cn/K_2^2}$$

with an absolute constant $C > 0$.

Proof. By Theorem 2.23 (see (2.23)) we have

$$(\text{vol}(B_{\Delta Q_n}(L_\infty)))^{1/N} \leq C(2^n n^2)^{-1/2} \leq Cn^{-1/2} N^{-1/2}$$

with an absolute constant $C > 0$. Using Theorem 2.24 we obtain

$$|\Omega| \geq 2|\Delta Q_n| e^{Cn/K_2^2}.$$

This proves Theorem 2.25. ■

Remark 2.26. In a particular case $K_2 = bn^\alpha$, $0 \leq \alpha \leq 1/2$, Theorem 2.25 gives

$$|\Omega| \geq 2|\Delta Q_n|e^{Cb^{-2}n^{1-2\alpha}}.$$

Corollary 2.27. Let a set $\Omega \subset \mathbb{T}^d$ have a property:

$$\forall t \in \mathcal{T}(\Delta Q_n) \quad \|t\|_\infty \leq bn^\alpha \|t\|_{\infty, \Omega}$$

with some $0 \leq \alpha < 1/2$. Then

$$|\Omega| \geq C_3 2^n n e^{Cb^{-2}n^{1-2\alpha}} \geq C_1(b, d, \alpha) |Q_n| e^{C_2(b, d, \alpha) n^{1-2\alpha}}.$$

Corollary 2.28. Let a set $\Omega \subset \mathbb{T}^2$ be such that $|\Omega| \leq C_5 |Q_n|$. Then

$$\sup_{f \in \mathcal{T}(Q_n)} \|f\|_\infty / \|f\|_{\infty, \Omega} \geq Cn^{1/2}.$$

Proof. Denote

$$K_2 := \sup_{f \in \mathcal{T}(Q_n)} \|f\|_\infty / \|f\|_{\infty, \Omega}.$$

Then the condition (2.24) of Theorem 2.25 is satisfied with this K_2 . Therefore, by Theorem 2.25

$$2|\Delta Q_n|e^{Cn/K_2^2} \leq |\Omega| \leq C_5 |Q_n|.$$

This implies that

$$K_2 \gtrsim n^{1/2}.$$

■

Remark 2.29. One can derive from the known results on recovery of functions from the classes \mathbf{W}_∞^r (see [335], [351]) that for any n there is a set $\Omega_n \subset \mathbb{T}^d$ such that $|\Omega_n| \leq C|Q_n|$ and

$$\sup_{f \in \mathcal{T}(Q_n)} (\|f\|_\infty / \|f\|_{\infty, \Omega_n}) \lesssim n^{d-1}.$$

For further results in this direction we refer the reader to a very recent paper [375].

2.6 Riesz products and the Small Ball Inequality

We consider the special trigonometric polynomial, which falls into a category of Riesz products (see [329])

$$\Phi_m(\mathbf{x}) := \prod_{k=0}^m (1 + \cos 4^k x_1 \cos 4^{m-k} x_2).$$

The above polynomial was the first example of the hyperbolic cross Riesz products. Clearly, $\Phi_m(\mathbf{x}) \geq 0$. It is known that

$$\Phi_m(\mathbf{x}) = 1 + \sum_{k=0}^m \cos 4^k x_1 \cos 4^{m-k} x_2 + t_m(\mathbf{x}), \quad t_m \in \mathcal{T}^\perp(4^m). \quad (2.25)$$

In particular, relation (2.25) implies that

$$\|\Phi_m\|_1 = 1.$$

We now consider a more general Riesz product in the case $d = 2$ (see [354] and [357]). For any two given integers $a \geq 1$ and $0 \leq b < a$ denote $AP(a, b)$ the arithmetic progression of the form $al + b$, $l = 0, 1, \dots$. Set

$$H_n(a, b) := \{\mathbf{s} : \mathbf{s} \in \mathbb{N}_0^2, \quad |\mathbf{s}|_1 = n, \quad s_1, s_2 \geq a, \quad s_1 \in AP(a, b)\}.$$

It will be convenient for us to consider subspaces $\mathcal{T}(\varrho'(\mathbf{s}))$ of trigonometric polynomials with harmonics in

$$\varrho'(\mathbf{s}) := \{\mathbf{k} \in \mathbb{Z}^2 : [2^{s_j-2}] \leq |k_j| < 2^{s_j}, \quad j = 1, 2\}.$$

For a subspace $Y \subset L_2(\mathbb{T}^2)$ we denote Y^\perp its orthogonal complement.

Lemma 2.30. *Take any trigonometric polynomials $t_{\mathbf{s}} \in \mathcal{T}(\varrho'(\mathbf{s}))$ and form the Riesz product*

$$\Phi(n, \mathbf{x}) := \prod_{\mathbf{s} \in H_n(a, b)} (1 + t_{\mathbf{s}}(\mathbf{x})).$$

Then for any $a \geq 6$ and any $0 \leq b < a$ this function admits the representation

$$\Phi(n, \mathbf{x}) = 1 + \sum_{\mathbf{s} \in H_n(a, b)} t_{\mathbf{s}}(\mathbf{x}) + R(\mathbf{x})$$

with $R \in \mathcal{T}(Q_{n+a-6})^\perp$.

Usually, Lemma 2.30 is used for $t_{\mathbf{s}}(\mathbf{x})$ being real trigonometric polynomials with $\|t_{\mathbf{s}}(\mathbf{x})\|_\infty \leq 1$.

The above Riesz products are useful in proving Small Ball Inequalities. We describe these inequalities for the Haar and the trigonometric systems. We begin with formulating this inequality in the case $d = 2$ using the dyadic enumeration of the Haar system

$$H_I(\mathbf{x}) := H_{I_1}(x_1)H_{I_2}(x_2), \quad \mathbf{x} = (x_1, x_2), \quad I = I_1 \times I_2.$$

Talagrand's inequality claims that for any coefficients $\{c_I\}$ (see [313] and [355])

$$\begin{aligned} \left\| \sum_{I: |I|=2^{-n}} c_I H_I(\mathbf{x}) \right\|_\infty &\geq 2^{-(n+1)} \sum_{I: |I|=2^{-n}} |c_I| \\ &= \frac{1}{2} \sum_{m=0}^n \left\| \sum_{I: |I_1|=2^{-m}, |I_2|=2^{m-n}} c_I H_I(\mathbf{x}) \right\|_1. \end{aligned} \quad (2.26)$$

where $|I|$ means the measure of I .

We now formulate an analogue of (2.26) for the trigonometric system. For an even number n define

$$Y_n := \{\mathbf{s} = (2n_1, 2n_2), \quad n_1 + n_2 = n/2\}.$$

Then for any coefficients $\{c_{\mathbf{k}}\}$ (see [354])

$$\left\| \sum_{\mathbf{s} \in Y_n} \sum_{\mathbf{k} \in \varrho(\mathbf{s})} c_{\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \right\|_\infty \geq C \sum_{\mathbf{s} \in Y_n} \left\| \sum_{\mathbf{k} \in \varrho(\mathbf{s})} c_{\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \right\|_1, \quad (2.27)$$

where C is a positive number. Inequality (2.27) plays a key role in the proof of lower bounds for the entropy numbers.

We proceed to the d -dimensional version of (2.27), which we formulate below. For even n , put

$$Y_n^d := \{\mathbf{s} = (2l_1, \dots, 2l_d), l_1 + \dots + l_d = n/2, l_j \in \mathbb{N}_0, j = 1, \dots, d\}.$$

It is conjectured (see, for instance, [200]) that the following inequality holds for any coefficients $\{c_{\mathbf{k}}\}$

$$n^{(d-2)/2} \left\| \sum_{\mathbf{s} \in Y_n^d} \sum_{\mathbf{k} \in \varrho(\mathbf{s})} c_{\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \right\|_{\infty} \geq C(d) \sum_{\mathbf{s} \in Y_n^d} \left\| \sum_{\mathbf{k} \in \varrho(\mathbf{s})} c_{\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \right\|_1. \quad (2.28)$$

We note that a weaker version of (2.28) with exponent $(d-2)/2$ replaced by $(d-1)/2$ is a direct corollary of the Parseval's identity, the Cauchy inequality and monotonicity of the L_p norms.

The d -dimensional version of (2.26), similar to the conjecture in (2.28), reads as follows:

$$\begin{aligned} n^{(d-2)/2} \left\| \sum_{I: |I|=2^{-n}} c_I H_I(\mathbf{x}) \right\|_{\infty} &\geq C(d) \sum_{\mathbf{s}: s_1 + \dots + s_d = n} \left\| \sum_{I: |I_j|=2^{-s_j}, j=1, \dots, d} c_I H_I(\mathbf{x}) \right\|_1 \\ &= C(d) 2^{-n} \sum_{I: |I|=2^{-n}} |c_I|. \end{aligned} \quad (2.29)$$

Recently, the authors of [36] and [37] proved (2.29) with the exponent $(d-1)/2 - \delta(d)$ with some $0 < \delta(d) < 1/2$ instead of $(d-2)/2$. See also its implications for Kolmogorov and entropy numbers of the mixed smoothness function classes in L_{∞} below and Subsection 6.4. Note, that there is no progress in proving (2.28).

2.7 Comments and open problems

Sections 2.1 and 2.2 mostly contain classical results on univariate trigonometric polynomials and their straight forward generalizations to the case of multivariate trigonometric polynomials with frequencies from parallelepipeds. Theorem 2.3 is from [344]. Its proof is based on the volume estimates from Theorem 2.18 and the classical Brun theorem on sections of convex bodies.

Lemma 2.4 is a direct corollary of Lemma 1.4 from [330] and the Littlewood-Paley theorem (see Appendix, Corollary 11.8). Another proof of Lemma 2.4 was given in [133]. It is easy to derive Lemma 2.4 from Theorem 2.11 (see [352], Ch.3). For Lemma 2.5 see [352], Ch.3.

Theorem 2.6 is the classical Bernstein inequality for the univariate trigonometric polynomials. Theorem 2.7 is a straight forward generalization of Theorem 2.6. Theorem 2.8 is discussed above.

Open problem 2.1. Find the order of the quantity

$$\sup_{t \in \mathcal{T}(N)} \|t^{(r)}(\mathbf{x}, \alpha)\|_1 / \|t\|_1$$

as a function on N .

Open problem 2.2. Find the order of the quantity

$$\sup_{t \in \mathcal{T}(N)} \sum_{\mathbf{k} \in \Gamma(N)} |\hat{t}(\mathbf{k})| / \|t\|_1$$

as a function on N .

Theorems 2.9 and 2.10 are classical Nikol'skii inequalities. Theorem 2.11, obtained in [340], is an important tool in hyperbolic cross approximation. Its proof in [340] is based on a nontrivial application of Theorem 2.10 and the Hölder inequalities (11.2). Theorems 2.13 and 2.14 are from [330] (see also [340]). Theorems 2.15 and 2.16 are classical variants of the Marcinkiewicz theorem. Theorem 2.17 is from [341].

Historical comments on results from Subsection 2.5 are given in the text above. We only formulate open problems in this regard.

Open problem 2.3. Find the order of

$$(\text{vol}(B_{\Delta Q_n}(L_\infty)))^{1/N}, \quad N := 2|\Delta Q_n|,$$

as $n \rightarrow \infty$, for $d \geq 3$.

Open problem 2.4. Find the order of

$$(\text{vol}(B_{\Delta Q_n}^\perp(L_1)))^{1/N}, \quad N := 2|\Delta Q_n|,$$

as $n \rightarrow \infty$, for $d \geq 3$.

Here are two fundamental open problems in connection with Subsection 2.6. Note, that Lemma 2.30 from [357] plays the key role in the two-dimensional version of (2.28).

Open problem 2.5. Prove the Small Ball Inequality (2.28) for $d \geq 3$.

Open problem 2.6. Prove the Small Ball Inequality (2.29) for $d \geq 3$.

3 Function spaces on \mathbb{T}^d

3.1 Spaces of functions with bounded mixed derivative

We begin with the univariate case in order to illustrate the action of the differential operator on periodic functions. For a trigonometric polynomial $f \in \mathcal{T}(n)$ we have

$$(D_x f)(x) = f'(x) = \sum_{|k| \leq n} ik \hat{f}(k) e^{ikx}.$$

We loose the information of $\hat{f}(0)$ when we differentiate. We can recover f from f' by the following formula

$$f(x) = \hat{f}(0) + f'(x) * \sum_{0 < |k| \leq n} (ik)^{-1} e^{ikx}.$$

Note that

$$\sum_{0 < |k| \leq n} (ik)^{-1} e^{ikx} = 2 \sum_{k=1}^n k^{-1} \sin kx.$$

Therefore, the following two definitions of the class W_p^1 , $1 \leq p \leq \infty$, are equivalent

$$(D1) \{f : \|f\|_p \leq 1, \|f'\|_p \leq 1\};$$

$$(D2) \{f : f = \varphi * F_1, \|\varphi\|_p \leq 1\},$$

where

$$F_1(x) := 1 + 2 \sum_{k=1}^{\infty} k^{-1} \sin kx = 1 + 2 \sum_{k=1}^{\infty} k^{-1} \cos(kx - \pi/2).$$

The second definition is more convenient than the first one for the following two reasons. It is easy to generalize it to the case of fractional (Weil) derivatives and it is easy to extend it to the multivariate case. We now give the general definition, which we use in this survey. This definition is based on the integral representation of a function by the Bernoulli kernels. Define for $x \in \mathbb{T}$ the univariate Bernoulli kernel

$$F_{r,\alpha}(x) := 1 + 2 \sum_{k=1}^{\infty} k^{-r} \cos(kx - \alpha\pi/2)$$

and define the multivariate Bernoulli kernels as the corresponding tensor products

$$F_{r,\alpha}(\mathbf{x}) := \prod_{j=1}^d F_{r,\alpha_j}(x_j), \quad \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{T}^d, \quad \alpha = (\alpha_1, \dots, \alpha_d). \quad (3.1)$$

Definition 3.1. Let $r > 0$, $\alpha \in \mathbb{R}$ and $1 \leq p \leq \infty$. Then $\mathbf{W}_{p,\alpha}^r$ is defined as the normed space of all $f \in L_p(\mathbb{T}^d)$ such that

$$f = F_{r,\alpha} * \varphi := (2\pi)^{-d} \int_{\mathbb{T}^d} F_{r,\alpha}(\mathbf{x} - \mathbf{y}) \varphi(\mathbf{y}) d\mathbf{y}$$

for some $\varphi \in L_p(\mathbb{T}^d)$, equipped with the norm $\|f\|_{\mathbf{W}_{p,\alpha}^r} := \|\varphi\|_p$.

It is well known and easy to prove that for all $r > 0$ and $\alpha \in \mathbb{R}$ we have $F_{r,\alpha} \in L_1$ (see [352], Ch. 1, Theorem 3.1). The extra parameter α allows us to treat simultaneously classes of functions with bounded mixed derivative and classes of functions with bounded trigonometric conjugate of the mixed derivative. In the case $\alpha = r$, Definition 3.1 is equivalent to the

mentioned below generalization of the definition (3.2)–(3.3) in terms of the Weil fractional derivatives (3.4). In the case $1 < p < \infty$, the parameter α does not play any role because the corresponding classes with different α are equivalent. In the case of classical derivative with natural r we set $\alpha = r$. In the case $\alpha = (r, \dots, r)$ we drop it from the notation: $\mathbf{W}_p^r := \mathbf{W}_{p,(r,\dots,r)}^r$. For simplicity of notations we formulate the majority of our results for classes \mathbf{W}_p^r . In those cases, when α affects the result we point it out explicitly.

We note that in the case $r \in \mathbb{N}$ and $1 \leq p \leq \infty$ the above class (space) can be described in terms of mixed partial derivatives. The Sobolev space \mathbf{W}_p^r of dominating mixed smoothness of order r can be defined as the collection of all $f \in L_p(\mathbb{T}^d)$ such that

$$D^{\mathbf{r}(e)} f \in L_p(\mathbb{T}^d), \quad \forall e \subset \{1, \dots, d\}, \quad (3.2)$$

where $\mathbf{r}(e)$ denotes the vector with components $r(e)_i = r$ for $i \in e$ and $r(e)_i = 0$ for $i \notin e$. Derivatives have to be understood in the weak sense. We endow this space with the norm

$$\|f\|_{\mathbf{W}_p^r(\mathbb{T}^d)} := \sum_{e \subset \{1, \dots, d\}} \|f\|_{\mathbf{W}_p^{\mathbf{r}(e)}}, \quad \|f\|_{\mathbf{W}_p^{\mathbf{r}(e)}} := \|D^{\mathbf{r}(e)} f\|_p. \quad (3.3)$$

This definition can be generalized to arbitrary $r \in \mathbb{R}$ based the $\mathbf{r}(e)$ -Weil fractional derivatives in the weak sense

$$D^{\mathbf{r}(e)} f := \sum_{\mathbf{k} \in \mathbb{Z}^d(e)} \left(\prod_{j \in e} (ik_j)^r \right) \hat{f}(\mathbf{k}) e^{i(\mathbf{k}, \cdot)}, \quad (3.4)$$

where $(ia)^b := |a|^b e^{i\pi b \operatorname{sign} a/2}$ for $a, b \in \mathbb{R}$, and $\mathbb{Z}^d(e) := \{\mathbf{k} \in \mathbb{Z}^d : k_j \neq 0, j \in e\}$ (see, e.g., [95, 108]).

For general $r > 0$ and $1 < p < \infty$ one may also use the condition

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{f}(\mathbf{k}) \prod_{j=1}^d (\max\{1, |k_j|\})^r e^{i(\mathbf{k}, \cdot)} \in L_p(\mathbb{T}^d).$$

In case $r \in \mathbb{N}$ this leads to an equivalent characterization.

By the Littlewood-Paley theorem, see Theorem 11.7 in the Appendix, we obtain a useful equivalent norm for \mathbf{W}_p^r in case $1 < p < \infty$ and $r > 0$, namely

$$\|f\|_{\mathbf{W}_p^r} \asymp \left\| \left(\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{r|\mathbf{s}|_1} |\delta_{\mathbf{s}}(f)(x)|^2 \right)^{1/2} \right\|_p. \quad (3.5)$$

This norm is particularly useful for the analysis of the approximation from the step-hyperbolic cross Q_n , see (2.15) above and Section 4 below.

3.2 Spaces of functions with bounded mixed difference

Let us first recall the basic concepts. For the univariate functions $f : \mathbb{T} \rightarrow \mathbb{C}$ the m th difference operator Δ_h^m is defined by

$$\Delta_h^m(f, x) := \sum_{j=0}^m (-1)^{m-j} \binom{m}{j} f(x + jh) \quad , \quad x \in \mathbb{T}, h \in [0, 1].$$

Let e be any subset of $\{1, \dots, d\}$. For multivariate functions $f : \mathbb{T}^d \rightarrow \mathbb{C}$ and $h \in [0, 1]^d$ the mixed (m, e) th difference operator $\Delta_h^{m,e}$ is defined by

$$\Delta_h^{m,e} := \prod_{i \in e} \Delta_{h_i}^m \quad \text{and} \quad \Delta_h^{m, \emptyset} = \operatorname{Id},$$

where $\text{Id } f = f$ and $\Delta_{h_i, i}^m$ is the univariate operator applied to the i -th variable of f with the other variables kept fixed. Let us refer to the recent survey [264] for general properties of mixed moduli of smoothness in L_p .

We first introduce spaces/classes \mathbf{H}_p^r of functions with bounded mixed difference.

Definition 3.2. Let $r > 0$ and $1 \leq p \leq \infty$. Fixing an integer $m > r$, we define the space \mathbf{H}_p^r as the set of all $f \in L_p(\mathbb{T}^d)$ such that for any $e \subset \{1, \dots, d\}$

$$\|\Delta_{\mathbf{h}}^{m,e}(f, \cdot)\|_p \leq C \prod_{i \in e} |h_i|^r$$

for some positive constant C , and introduce the norm in this space

$$\|f\|_{\mathbf{H}_p^r} := \sum_{e \subset \{1, \dots, d\}} |f|_{\mathbf{H}_p^r(e)},$$

where

$$|f|_{\mathbf{H}_p^r(e)} := \sup_{0 < |h_i| \leq 2\pi, i \in e} \left(\prod_{i \in e} |h_i|^{-r} \right) \|\Delta_{\mathbf{h}}^{m,e}(f, \cdot)\|_p.$$

Remark 3.3. Let us define the mixed (m, e) th modulus of smoothness by

$$\omega_m^e(f, \mathbf{t})_p := \sup_{|h_i| < t_i, i \in e} \|\Delta_{\mathbf{h}}^{m,e}(f, \cdot)\|_p, \quad \mathbf{t} \in [0, 1]^d, \quad (3.6)$$

for $f \in L_p(\mathbb{T}^d)$ (in particular, $\omega_m^\emptyset(f, t)_p = \|f\|_p$). Then there holds the following relation

$$|f|_{\mathbf{H}_p^r(e)} \asymp \sup_{2\pi > t_i > 0, i \in e} \left(\prod_{i \in e} t_i^{-r} \right) \omega_m^e(f, \mathbf{t})_p.$$

Based on this remark, we will introduce Besov spaces of mixed smoothness $\mathbf{B}_{p,\theta}^r$, a generalization of \mathbf{H}_p^r , see [4, 287, 396, 390].

Definition 3.4. Let $r > 0$ and $1 \leq p \leq \infty$. Fixing an integer $m > r$, we define the space $\mathbf{B}_{p,\theta}^r$ as the set of all $f \in L_p(\mathbb{T}^d)$ such that the norm

$$\|f\|_{\mathbf{B}_{p,\theta}^r} := \sum_{e \subset \{1, \dots, d\}} |f|_{\mathbf{B}_{p,\theta}^r(e)}$$

is finite, where for $e \subset \{1, \dots, d\}$

$$|f|_{\mathbf{B}_{p,\theta}^r(e)} := \begin{cases} \sup_{2\pi > t_i > 0, i \in e} \left(\prod_{i \in e} t_i^{-r} \right) \omega_m^e(f, \mathbf{t})_p, & \theta = \infty; \\ \left(\int_{(0, 2\pi)^d} \omega_m^e(f, \mathbf{t})_p^\theta \left(\prod_{i \in e} t_i^{-\theta r - 1} \right) d\mathbf{t} \right)^{1/\theta}, & \theta < \infty. \end{cases}$$

With this definition we have $\mathbf{B}_{p,\infty}^r = \mathbf{H}_p^r$. Notice that the definitions of $\mathbf{B}_{p,\theta}^r$ and \mathbf{H}_p^r are independent of m in the sense that different values of m induce equivalent quasi-norms of these spaces. With a little abuse of notation, denote the corresponding unit ball

$$\mathbf{B}_{p,\theta}^r := \{f : \|f\|_{\mathbf{B}_{p,\theta}^r} \leq 1\}.$$

Remark 3.5. *In many papers on hyperbolic cross approximation, especially from the former Soviet Union, instead of the spaces \mathbf{W}_p^r , \mathbf{H}_p^r and $\mathbf{B}_{p,\theta}^r$, the authors considered their subspaces. Namely, they studied functions f in \mathbf{W}_p^r , \mathbf{H}_p^r and $\mathbf{B}_{p,\theta}^r$, which satisfy an extra condition: f has zero mean value in each variable x_i , $i = 1, \dots, d$, that is,*

$$\int_{\mathbb{T}} f(\mathbf{x}) \, dx_i = 0.$$

However, this does not affect generality from the point of view of multivariate approximation (but not high-dimensional approximation, when we want to control dependence on dimension d) due to the following observation. Let \mathbf{F}_d temporarily denote one of the above spaces in d -variables. Then we have the following ANOVA-like decomposition for any $f \in \mathbf{F}_d$

$$f = c + \sum_{e \neq \emptyset} f_e,$$

where f_e are functions of $|e|$ variables x_i , $i \in e$, with zero mean values in the variables x_i , which can be treated as an element from $\mathbf{F}_{|e|}$. For details and bibliography see, [85, 88, 340, 352].

Remark 3.6. *It was understood in the beginning of the 1960s that hyperbolic crosses are closely related with the approximation and numerical integration of functions with dominating mixed smoothness which initiated a systematic study of these function classes. The following references have to be mentioned in connection with the development of the theory of function spaces with dominating mixed smoothness: Nikol'skii [242, 243], Babenko [7], Bakhvalov [15], Amanov [4], Temlyakov [340, 352], Tikhomirov [378, 379], Schmeißer, Triebel [287], Vybíral [396] and Triebel [382].*

3.3 Characterization via Fourier transform

In this subsection, we will give a characterization of spaces \mathbf{H}_p^r and $\mathbf{B}_{p,\theta}^r$ via Fourier transform. Let us first comment on the classical Korobov space \mathbf{E}_d^r introduced in [202], see also Subsection 9.1 below. For $r > 0$ we define the Korobov space

$$\mathbf{E}_d^r := \left\{ f \in L_1(\mathbb{T}^d) : \sup_{\mathbf{k} \in \mathbb{Z}^d} |\hat{f}(\mathbf{k})| \cdot \prod_{i=1}^d \max\{1, |k_i|\}^r < \infty \right\}.$$

The function $F_{r,\alpha}$ defined in (3.1) clearly belongs to \mathbf{E}_d^r . Using the Abel transformation twice, see Appendix 11.2, we can prove that for $\mathbf{s} \in \mathbb{N}_0^d$ it holds $\|A_{\mathbf{s}}(F_{r,\alpha})\|_1 \asymp 2^{-|\mathbf{s}|_1 r}$. In other words,

$$\sup_{\mathbf{s} \in \mathbb{N}_0^d} 2^{r|\mathbf{s}|_1} \|A_{\mathbf{s}}(F_{r,\alpha})\|_1 < \infty. \quad (3.7)$$

This immediately ensures that $F_{r,\alpha} \in L_1(\mathbb{T}^d)$. Moreover, (3.7) is exactly the condition for $F_{r,\alpha}$ belonging to \mathbf{H}_1^r as we will see below. In this sense, the classical Korobov space \mathbf{E}_d^r is slightly larger than the space \mathbf{H}_1^r . In Subsection 3.2 we have seen that Besov spaces $\mathbf{B}_{p,\theta}^r$ are defined in a classical way by using exclusively information on the “time side”, i.e., without any information on the Fourier coefficients. Such a useful tool is so far not available for the classical Korobov space \mathbf{E}_d^r .

Let us now characterize spaces \mathbf{H}_p^r and $\mathbf{B}_{p,\theta}^r$ via dyadic decompositions of the Fourier transform. We begin with the simplest version in terms of $\delta_{\mathbf{s}}(f)$. It is known that for $r > 0$ and $1 < p < \infty$,

$$\|f\|_{\mathbf{H}_p^r} \asymp \sup_{\mathbf{s}} \|\delta_{\mathbf{s}}(f)\|_p 2^{r|\mathbf{s}|_1}, \quad (3.8)$$

and

$$\|f\|_{\mathbf{B}_{p,\theta}^r} \asymp \left(\sum_{\mathbf{s}} \left(\|\delta_{\mathbf{s}}(f)\|_p 2^{r|\mathbf{s}|_1} \right)^\theta \right)^{1/\theta}, \quad (3.9)$$

and that for $r > 0$ and $1 \leq p \leq \infty$,

$$\|f\|_{\mathbf{H}_p^r} \asymp \sup_{\mathbf{s}} \|A_{\mathbf{s}}(f)\|_p 2^{r|\mathbf{s}|_1},$$

and

$$\|f\|_{\mathbf{B}_{p,\theta}^r} \asymp \left(\sum_{\mathbf{s}} \left(\|A_{\mathbf{s}}(f)\|_p 2^{r|\mathbf{s}|_1} \right)^\theta \right)^{1/\theta}$$

(see, e.g., [88, 95, 340, 352, 287, 390, 239]).

The characterizations in the right hand side of (3.8) (3.9) are simple and work well for $1 < p < \infty$. In the cases $p = 1$ and $p = \infty$ the operators $\delta_{\mathbf{s}}(\cdot)$ are not uniformly bounded as operators from L_p to L_p . This issue is resolved by replacing operators $\delta_{\mathbf{s}}(\cdot)$ by operators $A_{\mathbf{s}}(\cdot)$. Such a modification gives equivalent definitions of classes $\mathbf{B}_{p,\theta}^r$ in the case $1 \leq p \leq \infty$. We now present a general way for characterizing the Besov classes for $0 < p \leq 1$ in the spirit as done in [287, Chapt. 2]. In order to proceed to $0 < p \leq 1$ we need the concept of a smooth dyadic decomposition of unity.

Definition 3.7. Let $\Phi(\mathbb{R})$ be the class of all systems $\varphi = \{\varphi_n(x)\}_{n=0}^\infty \subset C_0^\infty(\mathbb{R})$ satisfying

- (i) $\text{supp } \varphi_0 \subset \{x : |x| \leq 2\}$,
- (ii) $\text{supp } \varphi_n \subset \{x : 2^{n-1} \leq |x| \leq 2^{n+1}\}$, $n = 1, 2, \dots$,
- (iii) For all $\ell \in \mathbb{N}_0$ it holds $\sup_{x,n} 2^{n\ell} |D^\ell \varphi_n(x)| \leq c_\ell < \infty$,
- (iv) $\sum_{n=0}^\infty \varphi_n(x) = 1$ for all $x \in \mathbb{R}$.

Remark 3.8. The class $\Phi(\mathbb{R})$ is not empty. We consider the following standard example. Let $\varphi_0(x) \in C_0^\infty(\mathbb{R})$ be a smooth function with $\varphi_0(x) = 1$ on $[-1, 1]$ and $\varphi_0(x) = 0$ if $|x| > 2$. For $n > 0$ we define

$$\varphi_n(x) = \varphi_0(2^{-n}x) - \varphi_0(2^{-n+1}x).$$

It is easy to verify that the system $\varphi = \{\varphi_n(x)\}_{n=0}^\infty$ satisfies (i) - (iv).

Now we fix a system $\{\varphi_n\}_{n \in \mathbb{Z}} \in \Phi(\mathbb{R})$, where we put $\varphi_n \equiv 0$ if $n < 0$. For $\mathbf{s} = (s_1, \dots, s_d) \in \mathbb{Z}^d$ let the building blocks $f_{\mathbf{s}}$ be given by

$$f_{\mathbf{s}}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \varphi_{s_1}(k_1) \cdot \dots \cdot \varphi_{s_d}(k_d) \hat{f}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}} \quad , \quad \mathbf{x} \in \mathbb{T}^d, \mathbf{s} \in \mathbb{N}_0^d. \quad (3.10)$$

Definition 3.9. Let $0 < p, \theta \leq \infty$ and $r > \sigma_p := (1/p - 1)_+$. Then $\mathbf{B}_{p,\theta}^r$ is defined as the collection of all $f \in L_1(\mathbb{T}^d)$ such that

$$\|f\|_{\mathbf{B}_{p,\theta}^r}^\varphi := \left(\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{|\mathbf{s}|_1 r \theta} \|f_{\mathbf{s}}\|_p^\theta \right)^{1/\theta}$$

is finite (usual modification in case $q = \infty$).

Recall, that this definition is independent of the chosen system φ in the sense of equivalent (quasi-)norms. Moreover, in case $\min\{p, q\} \geq 1$ the defined spaces are Banach spaces, whereas they are quasi-Banach spaces in case $\min\{p, q\} < 1$. For details confer [287, 2.2.4].

As already mentioned above the two approaches for the definition of the Besov spaces $\mathbf{B}_{p,\theta}^r$ of mixed smoothness are equivalent if $p, \theta \geq 1$. Concerning difference characterizations for the quasi-Banach range of parameters there are still some open questions, see [287, 2.3.4, Rem. 2] and Theorem 9.5 below. Let us state the following general equivalence result.

Lemma 3.10. *Let $1 \leq p, \theta \leq \infty$ and $m \in \mathbb{N}$ with $m > r > 0$. Then*

$$\|f\|_{\mathbf{B}_{p,\theta}^r}^\varphi \asymp \|f\|_{\mathbf{B}_{p,\theta}^r}, \quad f \in L_1(\mathbb{T}^d).$$

As already mentioned above we have the equivalent characterization (3.5) for the spaces \mathbf{W}_p^r in case $r > 0$ and $1 < p < \infty$. There is also a characterization in terms of the so-called rectangular means of differences, i.e.,

$$\mathcal{R}_m^e(f, \mathbf{t}, \mathbf{x}) := \frac{1}{t_1} \int_{[-t_1, t_1]} \cdots \frac{1}{t_d} \int_{[-t_d, t_d]} |\Delta_{(h_1, \dots, h_d)}^{m,e}(f, \mathbf{x})| dh_d \dots dh_1, \quad \mathbf{x} \in \mathbb{R}^d, \mathbf{t} \in (0, 1]^d. \quad (3.11)$$

The following lemma is a straight-forward modification of [390, Thm. 3.4.1], see also [239].

Lemma 3.11. *Let $1 < p < \infty$ and $r > 0$. Let further $m \in \mathbb{N}$ be a natural number with $m > r$. Then*

$$\|f\|_{\mathbf{W}_p^r} \asymp \|f\|_{\mathbf{W}_p^r}^{(m)}, \quad f \in L_1(\mathbb{R}),$$

where

$$\|f\|_{\mathbf{W}_p^r}^{(m)} := \left\| \left(\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{r|\mathbf{s}|_1} \mathcal{R}_m^{e(\mathbf{s})}(f, 2^{-\mathbf{s}}, \cdot)^2 \right)^{1/2} \right\|_p$$

and $2^{-\mathbf{s}} := (2^{-s_1}, \dots, 2^{-s_d})$.

3.4 Embeddings

Here we review some useful embeddings between the classes $\mathbf{B}_{p,\theta}^r$ and \mathbf{W}_p^r .

Lemma 3.12. *Let $0 < p < u \leq \infty$, $\beta = 1/p - 1/u$, $r \in \mathbb{R}$, and $0 < \theta \leq \infty$.*

(i) *It holds*

$$\mathbf{B}_{p,\theta}^r \hookrightarrow \mathbf{B}_{u,\theta}^{r-\beta}.$$

(ii) *If in addition $1 < p < u < \infty$ then*

$$\mathbf{W}_p^r \hookrightarrow \mathbf{W}_u^{r-\beta}.$$

Both embeddings are non-compact.

(iii) *If $r > 1/p$ then the embedding*

$$\mathbf{B}_{p,\theta}^r \hookrightarrow C(\mathbb{T}^d) \quad (3.12)$$

is compact. In case $r = 1/p$ and $\theta \leq 1$ the embedding (3.12) keeps valid but is not compact.

(iv) *If $1 < p < \infty$ and $r > 0$ then*

$$\mathbf{B}_{p,\min\{p,2\}}^r \hookrightarrow \mathbf{W}_p^r \hookrightarrow \mathbf{B}_{p,\max\{p,2\}}^r. \quad (3.13)$$

Let us particularly mention the following two non-trivial embeddings between \mathbf{B} and \mathbf{W} -spaces for different metrics.

Lemma 3.13. *Let $0 < p < u < \infty$, $1 < u < \infty$ and $\beta := 1/p - 1/u$. Then for $r \geq \beta$ we have*

$$\mathbf{B}_{p,u}^r \hookrightarrow \mathbf{W}_u^{r-\beta}. \quad (3.14)$$

Lemma 3.14. *Let $1 < p < u \leq \infty$ and $\beta := 1/p - 1/u$. Then for $r \geq \beta$ we have*

$$\mathbf{W}_p^r \hookrightarrow \mathbf{B}_{u,p}^{r-\beta}.$$

Remark 3.15. *Lemma 3.13 follows from Theorem 2.11 (see (2.20) with $p = u$, $q = p$). The embedding in Lemma 3.13 is nontrivial and very useful in analysis of approximation of classes with mixed smoothness. In the univariate case an analog of (3.14) was obtained by Ul'yanov [393] and Timan [380]. They used different methods of proof. Their techniques work for the multivariate case of isotropic Besov spaces as well. Franke [126] proved (3.14) for isotropic Besov spaces on \mathbb{R}^d and obtained its version with the \mathbf{W} space replaced by the appropriate Triebel-Lizorkin spaces. The converse embedding in Lemma 3.14 for isotropic spaces (a Triebel-Lizorkin space embedded in an appropriate Besov space) has been obtained by Jawerth [184]. Lemma 3.14 is a corollary of Theorem 2.11. It directly follows from Theorem 2.11 in the special case $r = 0$. The case $r > 0$ follows from the case $r = 0$ and the well known relation $\|t^{(r)}\|_p \asymp 2^{r|s|} \|t\|_p$ for $t \in \mathcal{T}(\rho(\mathbf{s}))$. A new proof of both relations based on atomic decompositions has been given recently by Vybíral [397]. The step from the univariate and isotropic multivariate cases to the case of mixed smoothness spaces required a modification of technique. In the periodic case it was done by Temlyakov [334], [340] (see Theorem 2.11 above) and in the case of \mathbb{R}^d by Hansen and Vybíral [167].*

Let us finally complement the discussion from the beginning of Section 3.3 and state useful embedding relations in the situation $p = 1$.

Lemma 3.16. *Let $r > 0$ and $\alpha \in \mathbb{R}$. Then the following continuous embeddings hold true.*

$$\mathbf{B}_{1,1}^r \hookrightarrow \mathbf{W}_{1,\alpha}^r \hookrightarrow \mathbf{B}_{1,\infty}^r \hookrightarrow \mathbf{E}_d^r. \quad (3.15)$$

We note that in case $p = 1$ we use operators $A_{\mathbf{s}}$ instead of $\delta_{\mathbf{s}}$ in the characterization of the \mathbf{B} classes. The first relation follows from Theorem 2.7. The second relation follows from (3.7) and the third embedding is a simple consequence of the characterization of $\mathbf{B}_{1,\infty}^r$ together with $|\hat{f}(\mathbf{k})| \leq \|f\|_1$.

Note, that the embedding $\mathbf{W}_{1,\alpha}^r \hookrightarrow \mathbf{B}_{1,2}^r$, as a formal counterpart of (3.13), does not hold true here. In fact, it does not even hold true with $\mathbf{B}_{1,\theta}^r$ and $2 < \theta < \infty$ on the right-hand side. In that sense, the embedding (3.15) is sharp. Note also, that the embeddings in Lemma 3.16 are strict. The (tensorized and) periodized hat function, see Figure 11 below, belongs to $\mathbf{B}_{1,\infty}^2$ but not to $\mathbf{W}_{1,\alpha}^2$.

4 Linear approximation

4.1 Introduction

By *linear approximation* we understand approximation from a fixed finite dimensional subspace. In the study of approximation of the univariate periodic functions the idea of representing a function by its Fourier series is very natural and traditional. It goes back to the work of Fourier from 1807. In this case one can use as a natural tool of approximation the partial sums of the Fourier expansion. In other words this means that we use the subspace $\mathcal{T}(n)$ for a source of approximants and use the orthogonal projection onto $\mathcal{T}(n)$ as the approximation operator. This natural approach is based on a standard ordering of the trigonometric system: $1, e^{ikx}, e^{-ikx}, e^{2ikx}, e^{-2ikx}, \dots$. We loose this natural approach, when we go from the univariate case to the multivariate case – there is no natural ordering of the \mathcal{T}^d for $d > 1$. The following idea of choosing appropriate trigonometric subspaces for approximation of a given class \mathbf{F} of multivariate functions was suggested by Babenko [6]. This idea is based on the concept of the Kolmogorov width introduced in [201]: for a centrally symmetric compact $\mathbf{F} \subset X$ define

$$d_m(\mathbf{F}, X) := \inf_{\varphi_1, \dots, \varphi_m} \sup_{f \in \mathbf{F}} \inf_{c_1, \dots, c_m} \|f - \sum_{k=1}^m c_k \varphi_k\|_X.$$

Consider a Hilbert space $L_2(\mathbb{T}^d)$ and suppose that the function class $\mathbf{F} = A(B(L_2))$ of our interest is an image of the unit ball $B(L_2)$ of $L_2(\mathbb{T}^d)$ under a mapping $A : L_2(\mathbb{T}^d) \rightarrow L_2(\mathbb{T}^d)$ of a compact operator A . For instance, in the case of $\mathbf{F} = \mathbf{W}_2^r$ the operator $A := A_r$ is the convolution with the kernel $F_r(\mathbf{x})$. It is now well known and was established by Babenko [6] for a special class of operators A that

$$d_m(A(B(L_2)), L_2) = s_{m+1}(A),$$

where $s_j(A)$ are the singular numbers of the operator A : $s_j(A) = (\lambda_j(AA^*))^{1/2}$.

Suppose now that the eigenfunctions of the operator AA^* are the trigonometric functions $e^{i(\mathbf{k}^j, \mathbf{x})}$. Then the optimal in the sense of the Kolmogorov width m -dimensional subspace will be the span $\{e^{i(\mathbf{k}^j, \mathbf{x})}\}_{j=1}^m$. Applying this approach to the class \mathbf{W}_2^r we obtain that for $m = |\Gamma(N)|$ the optimal subspace for approximation in L_2 is the subspace of hyperbolic cross polynomials $\mathcal{T}(\Gamma(N))$. This observation led to a thorough study of approximation by the hyperbolic cross polynomials. We discuss it in Subsection 4.2.

B.S. Mityagin [233] used the harmonic analysis technique, in particular, the Marcinkiewicz multipliers (see Theorem 11.10), to prove that

$$d_m(\mathbf{W}_p^r, L_p) \asymp m^{-r} (\log m)^{r(d-1)}, \quad 1 < p < \infty.$$

He also proved that optimal, in the sense of order, subspaces are $\mathcal{T}(Q_n)$ with $|Q_n| \asymp m$ and $|Q_n| \leq m$. In addition, the operator $S_{Q_n}(\cdot)$ of orthogonal projection onto $\mathcal{T}(Q_n)$ can be taken as an approximation operator. The use of harmonic analysis techniques for the L_p spaces lead to the change from smooth hyperbolic crosses $\Gamma(N)$ to step hyperbolic crosses Q_n . The idea of application of the theory of widths for finding good subspaces for approximation of classes of functions with mixed smoothness is very natural and was used in many papers. A typical problem here is to study approximation of classes \mathbf{W}_p^r in the L_q for all $1 \leq p, q \leq \infty$. We give a detailed discussion of these results in further subsections. We only give a brief qualitative remarks on those results in this subsection. As we mentioned above, in linear approximation we are interested in approximation from finite dimensional subspaces. The

Kolmogorov width provides a way to determine optimal (usually, in the sense of order) m -dimensional subspaces. The approximation operator, used in the Kolmogorov width, is the operator of best approximation. Clearly, we would like to use as simple approximation operators as possible. As a result the following widths were introduced and studied.

V.M. Tikhomirov [376] introduced the concept of linear width:

$$\lambda_m(\mathbf{F}, L_p) := \inf_{A: \text{rank } A \leq m} \sup_{f \in \mathbf{F}} \|f - A(f)\|_p.$$

V.N. Temlyakov [325] introduced the concept of orthowidth (Fourier width):

$$\varphi_m(\mathbf{F}, L_p) := \inf_{u_1, \dots, u_m} \sup_{f \in \mathbf{F}} \|f - \sum_{j=1}^m \langle f, u_j \rangle u_j\|_p,$$

where u_1, \dots, u_m is an orthonormal system.

It is clear that for any class \mathbf{F} and $1 \leq p \leq \infty$

$$d_m(\mathbf{F}, L_p) \leq \lambda_m(\mathbf{F}, L_p) \leq \varphi_m(\mathbf{F}, L_p) \quad (4.1)$$

and for $p = 2$

$$d_m(\mathbf{F}, L_2) = \lambda_m(\mathbf{F}, L_2) = \varphi_m(\mathbf{F}, L_2).$$

We now give a brief comparison of the above three widths in the case of classes W_p^r of univariate functions. For convenience we denote

$$D_1 := \{(p, q) : 1 \leq p \leq q \leq 2 \text{ or } 1 \leq q \leq p \leq \infty\},$$

$$D_2 := \{(p, q) : 2 \leq p < q \leq \infty\},$$

$$D_3 := \{(p, q) : 1 \leq p < 2 \text{ and } 2 < q \leq \infty\},$$

and let D_4 be a part of D_3 such that $1/q + 1/p \geq 1$ and $D_5 = D_3 \setminus D_4$. It is convenient to represent the corresponding domains in term of points $(1/p, 1/q)$ on the square $[0, 1]^2$ instead of points (p, q) on $[1, \infty]^2$. Denote

$$D_i^* := \{(1/p, 1/q) : (p, q) \in D_i\}, \quad i = 1, \dots, 5.$$

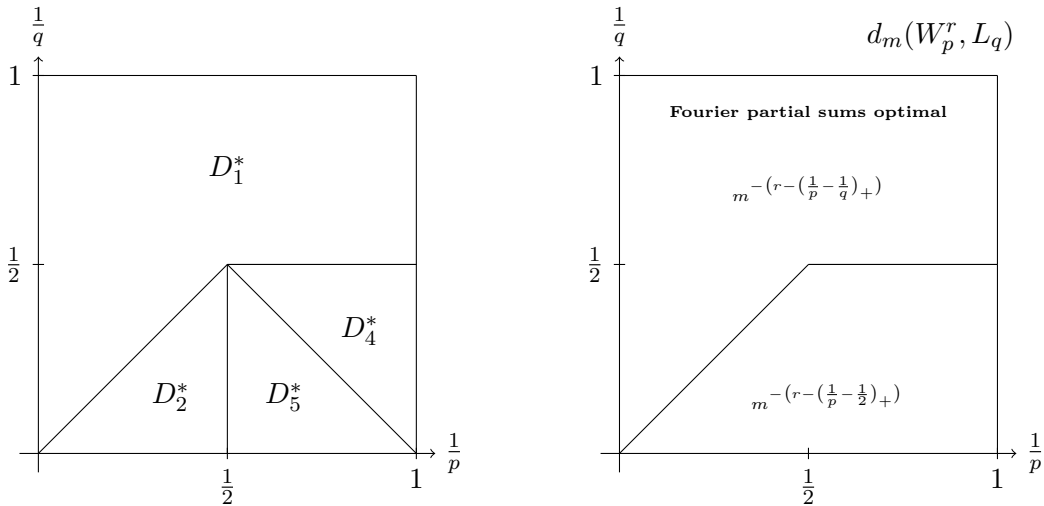


Figure 3: The univariate behavior of $d_m(W_p^r, L_q)$ in different regions

It is known that for $(p, q) \in D_1$ approximation by trigonometric polynomials in $\mathcal{T}(n)$ with $n = \lceil (m-1)/2 \rceil$ gives the order of decrease of the Kolmogorov widths $d_m(W_p^r, L_q)$. But for $(p, q) \notin D_1$ this is not the case:

$$d_{2n+1}(W_p^r, L_q) = o(E_n(W_p^r)_q), \quad (p, q) \notin D_1.$$

For $(p, q) \in D_1$ the orders of the Kolmogorov widths can be obtained by linear operators V_n , $m = 4n - 1$. It is known that the operators V_n give the orders of linear widths not only in the domain D_1 but also in the domain D_2 . In the domain D_3 the relation

$$\lambda_{2n+1}(W_p^r, L_q) = o(E_n(W_p^r)_q)$$

holds, which shows, in particular, that for $(p, q) \in D_3$ the order of linear width can not be realized by the operators V_n .

For $(p, q) \in D_1 \cup D_4$ the orders of the Kolmogorov widths can be realized by linear methods: in the case $(p, q) \in D_1$ by means of the operators V_n , and in the case $(p, q) \in D_4$ by means of some other linear operators.

For the classes W_p^r for all (p, q) , excepting the case $(p, q) = (1, 1), (\infty, \infty)$, the operators S_n ($m = 2n + 1$) are optimal Fourier operators in the sense of order.

The linear operators A providing the orders of the widths $\lambda_m(W_p^r, L_q)$ for $(p, q) \in D_3$, that is, in the case when A differs from V_n ($m = 4n - 1$), are not orthogonal projections. Moreover, operators A can not be bounded uniformly (over m) as operators from L_2 to L_2 .

Further, for example for $p = 2$ and $q = \infty$, the Kolmogorov widths decrease faster than the corresponding linear widths:

$$d_m(W_2^r, L_\infty) \asymp m^{-1/2} \lambda_m(W_2^r, L_\infty).$$

However, up to now no concrete example of a system $\{\varphi_i\}_{i=1}^m$ is known, the best approximations by which would give the order of $d_m(W_2^r, L_\infty)$ (the same is true for the domain $D_2 \cup D_5$).

This discussion shows that the sets $\mathcal{T}(n)$ and the operators V_n and S_n are optimal in many cases from the point of view of the Kolmogorov widths, linear widths and orthowidths. In the cases when we can approximate better than by means of the operators V_n and S_n , we must sacrifice some useful properties which these operators have.

We have a similar qualitative picture in the case of approximation of classes of functions with mixed smoothness. The role of $\mathcal{T}(n)$ is played now by $\mathcal{T}(Q_n)$ with $m \asymp |Q_n|$. The analog of the univariate de la Vallée Poussin kernel $\mathcal{V}_n(x)$ the kernel \mathcal{V}_{Q_n} (see Section 2) is not as good as its univariate version. Lemma 2.5 from Section 2 gives

$$\|\mathcal{V}_{Q_n}\|_1 \asymp n^{d-1}.$$

This substantially complicates the study of approximation in L_1 and L_∞ norms. Many problems of approximation in these spaces are still open. For $1 < p < \infty$ the role of S_n is played by S_{Q_n} .

4.2 Approximation by the hyperbolic cross polynomials

The operators S_{Q_n} and V_{Q_n} play an important role in the hyperbolic cross approximation. These operators can be written in terms of the corresponding univariate operators in the following form. Denote by S_l^i the univariate operator S_l acting on functions on the variable x_i . Then, it follows from the definition of Q_n that

$$S_{Q_n} = \sum_{\mathbf{s}: |\mathbf{s}|_1 \leq n} \prod_{i=1}^d (S_{2^{s_i-1}}^i - S_{[2^{s_i-1}]-1}^i), \quad (4.2)$$

here $S_{-1} = 0$. A similar formula holds for the V_{Q_n} .

The Smolyak algorithm

Operators of the form (4.2) with S^i replaced by other univariate operators are used in sampling recovery (see Section 5) and other problems. For a generic discussion we refer to Novak [247] and Wasilkowski, Woźniakowski [401]. Sometimes Smolyak's algorithm can also be identified in the framework of boolean methods, see [68].

The approximate recovery operators of the form (4.2) were first considered by Smolyak [305]. A standard name for operators of the form (4.2) is *Smolyak-type algorithms*. Very often analysis of operators S_{Q_n} , V_{Q_n} , and other operators of the form (4.2) goes along the same lines. The following general framework was suggested in [5], see also [298]. Let three numbers $a > b \geq 0$, and $1 \leq p \leq \infty$ be given. Consider a family of univariate linear operators $\{Y_s\}_{s=0}^\infty$, which are defined on the space W_p^a and have the following two properties:

- (1) For any f from the class W_p^a we have

$$\|f - Y_s(f)\|_p \leq C_1 2^{-as} \|f\|_{W_p^a}, \quad s = 0, 1, 2, \dots;$$

- (2) For any trigonometric polynomial t of order 2^u , we have

$$\|Y_s(t)\|_p \leq C_2 2^{b(u-s)} \|t\|_p, \quad u \geq s.$$

Let as above Y_s^i denote the univariate operator Y_s acting on functions on the variable x_i . Consider the following d -dimensional operator

$$T_n := \sum_{\mathbf{s}: |\mathbf{s}|_1 \leq n} \Delta_{\mathbf{s}}, \quad \Delta_{\mathbf{s}} := \prod_{i=1}^d (Y_{s_i}^i - Y_{s_i-1}^i),$$

with $Y_{-1} = 0$. We illustrate the above general setting by one result from [5].

Proposition 4.1. *Let operators $\{Y_s\}_{s=0}^\infty$ satisfy conditions (1) and (2). Then for any $r \in (b, a)$ we have for $f \in \mathbf{H}_p^r$*

$$\|\Delta_{\mathbf{s}}(f)\|_p \lesssim 2^{-r|\mathbf{s}|_1} \|f\|_{\mathbf{H}_p^r} \quad \text{and} \quad \|f - T_n(f)\|_p \lesssim 2^{-rn} n^{d-1} \|f\|_{\mathbf{H}_p^r}.$$

We note that the technique developed in [48], see Subsection 5.2, allows for extending Proposition 4.1 to classes $\mathbf{B}_{p,\theta}^r$.

Proposition 4.2. *Let operators $\{Y_s\}_{s=0}^\infty$ satisfy conditions (1) and (2). Then for any $r \in (b, a)$ we have for $f \in \mathbf{B}_{p,\theta}^r$*

$$\left(\sum_{\mathbf{s}} \left(2^{r|\mathbf{s}|_1} \|\Delta_{\mathbf{s}}(f)\|_p \right)^\theta \right)^{1/\theta} \lesssim \|f\|_{\mathbf{B}_{p,\theta}^r}$$

and

$$\|f - T_n(f)\|_p \lesssim 2^{-rn} n^{(d-1)(1-1/\theta)} \|f\|_{\mathbf{B}_{p,\theta}^r}.$$

Approximation from the hyperbolic cross

We now proceed to best approximation by the hyperbolic cross polynomials from $\mathcal{T}(Q_n)$. Denote for a function $f \in L_q$

$$E_{Q_n}(f)_q := \inf_{t \in \mathcal{T}(Q_n)} \|f - t\|_q$$

and for a function class \mathbf{F}

$$E_{Q_n}(\mathbf{F})_q := \sup_{f \in \mathbf{F}} E_{Q_n}(f)_q.$$

We begin with approximation of classes \mathbf{W}_p^r in L_q . It is clear that approximation of the Bernoulli kernels F_r , which are used for integral representation of a function from \mathbf{W}_p^r , plays an important role in approximation of classes \mathbf{W}_p^r . The following theorem is from [323]

Theorem 4.3. *For $1 \leq q \leq \infty$ and $r - 1 + 1/q > 0$ we have*

$$E_{Q_n}(F_r)_q \asymp 2^{-(r-1+1/q)n} n^{(d-1)/q}.$$

By the corollary to the Littlewood-Paley theorem we get from this estimate,

$$\|F_r - S_{Q_n}(F_r)\|_q \lesssim 2^{-(r-1+1/q)n} n^{(d-1)/q}, \quad 1 < q < \infty.$$

In the case of $q = \infty$ we only get

$$\|F_r - S_{Q_n}(F_r)\|_\infty \lesssim 2^{-n(r-1)} n^{d-1},$$

which has an extra factor n^{d-1} compared to $E_{Q_n}(F_r)_\infty$.

The upper bounds for the best approximations of the functions F_r in the uniform metric were obtained (see [330]) using the Nikol'skii duality theorem (see Appendix, Theorem 11.3. The use of the duality theorem has the result that we can determine the order of the best approximation of F_r in the uniform metric, but we cannot construct a polynomial giving this approximation. The situation is unusual from the point of view of approximation of functions of one variable.

Theorem 4.4. *Suppose that $1 < p, q < \infty$, $r > (1/p - 1/q)_+$. Then*

$$E_{Q_n}(\mathbf{W}_p^r)_q \asymp 2^{-n(r-(1/p-1/q)_+)}.$$

This theorem was proved by harmonic analysis technique for the case $p = q$ in [233], [241] and for $q \neq p$ in [131].

We now consider the cases when one or two parameters p, q take the extreme value 1 or ∞ . These results are not as complete as in the case $1 < p, q < \infty$.

Theorem 4.5. *We have*

$$E_{Q_n}(\mathbf{W}_p^r)_q \asymp \begin{cases} 2^{-nr}, & r > 0; \quad p = \infty, \quad 1 \leq q < \infty; \quad 1 < p \leq \infty, \quad q = 1; \\ 2^{-n(r-1+1/q)} n^{\frac{d-1}{q}}, & r > 1 - 1/q, \quad p = 1, \quad 1 \leq q \leq \infty; \\ 2^{-n(r-1/p)}, & r > 1/p, \quad 1 \leq p \leq 2, \quad q = \infty. \end{cases}$$

The upper bounds in the case $p = 1$ follow from Theorem 4.3. The case $1 \leq p \leq 2, q = \infty$, was established in [329]. For the first case see [352], Ch. 3, Theorem 3.4.

We now proceed to classes \mathbf{H}_p^r . The problem of finding the right orders of decay of the $E_{Q_n}(\mathbf{H}_p^r)_q$ turns out to be more difficult than the corresponding problem for the \mathbf{W}_p^r classes. Even in the case $1 < p, q < \infty$ a new technique was required.

Theorem 4.6. *We have*

$$E_{Q_n}(\mathbf{H}_p^r)_q \asymp \begin{cases} 2^{-n(r-1/p+1/q)} n^{\frac{d-1}{q}}, & 1 \leq p < q < \infty, \quad r > 1/p - 1/q, \\ 2^{-rn} n^{\frac{d-1}{2}}, & 1 < q \leq p < \infty, \quad p \geq 2; \quad p = \infty, \quad 1 < q < \infty, \\ 2^{-rn} n^{\frac{d-1}{p}}, & 1 \leq q \leq p \leq 2, \quad r > 0. \end{cases}$$

In the case $p = q = 2$ Theorem 4.6 was proved in [42] and in the case $1 < p = q < \infty$ in [241]. In the case $p = q = 1$ it was proved in [329] and in the case $1 \leq p < q \leq 2$ in [331]. In the case $1 \leq p < q < \infty$, Theorem 4.6 was proved in [340] with the use of Theorem 2.11 from Section 2. In the case $2 \leq q < p \leq \infty$ the required upper bounds follow from the upper bounds in the case $1 < q = p < \infty$ and the lower bounds follow from [324]. In the case $1 < q < 2 \leq p < \infty$ Theorem 4.6 was proved in [134] and [86]. In the case $1 \leq q < p \leq 2$ the proof of lower bounds required a new technique (see [339]).

Theorem 4.7. *Let $2 < p \leq \infty$ and $r > 0$. Then we have*

$$E_{Q_n}(\mathbf{H}_p^r)_1 \asymp 2^{-rn} n^{\frac{d-1}{2}}.$$

The upper bounds in Theorem 4.7 follow from the upper bounds for $E_{Q_n}(\mathbf{H}_2^r)_2$ from Theorem 4.6. The lower bounds are nontrivial. They follow from the corresponding lower bounds for the Kolmogorov widths $d_m(\mathbf{H}_\infty^r, L_1)$, which, as it was observed in [31], follow from the lower bounds for the entropy numbers $\epsilon_k(\mathbf{H}_\infty^r, L_1)$ from [342] and [344].

The following result is known in the case of functions of two variables.

Theorem 4.8. *Let $d = 2$, and $r > 0$. Then*

$$E_{Q_n}(\mathbf{H}_\infty^r)_q \asymp \begin{cases} 2^{-rn} n & \text{for } q = \infty, \\ 2^{-rn} n^{1/2} & \text{for } q = 1. \end{cases}$$

This theorem was proved in [329] in the case $q = \infty$ and in [352], Chapter 3, Theorem 3.5, in the case $q = 1$. The proof of lower bounds is based on the Riesz products (see Subsection 2.6).

In this subsection we studied approximation in the L_q -metric of functions in the classes \mathbf{W}_p^r and \mathbf{H}_p^r , $1 \leq p, q \leq \infty$ by trigonometric polynomials whose harmonics lie in the hyperbolic crosses. Certain specific features of the multidimensional case were observed in this study.

As is known, in the univariate case the order of the least upper bounds of the best approximation by trigonometric polynomials for both classes are the same for all $1 \leq p, q \leq \infty$, even though H_p^r is a wider class than W_p^r . It was determined that the least upper bounds of the best approximation by polynomials in $\mathcal{T}(Q_n)$ are different for the classes \mathbf{W}_p^r and \mathbf{H}_p^r for all $1 < p, q < \infty$. Namely,

$$E_{Q_n}(\mathbf{W}_p^r)_q = o(E_{Q_n}(\mathbf{H}_p^r))_q, \quad d \geq 2.$$

This phenomenon is related to the following fact. Let $1 < p < \infty$. The property $f \in \mathbf{W}_p^r$ implies and is very close to the property

$$\|f_l\|_p \lesssim 2^{-rl}, \quad f_l := \sum_{|\mathbf{s}|_1=l} \delta_{\mathbf{s}}(f).$$

Contrary to that the property $f \in \mathbf{H}_p^r$ is equivalent to

$$\|\delta_{\mathbf{s}}(f)\|_p \lesssim 2^{-r|\mathbf{s}|_1}.$$

Therefore, the difference between classes \mathbf{W}_p^r and \mathbf{H}_p^r is determined by the interplay between conditions on the dyadic blocks and the hyperbolic layers. The number of the dyadic blocks in the n th hyperbolic layer is of $\asymp n^{d-1}$. The quantities $E_{Q_n}(\mathbf{W}_p^r)_q$ and $E_{Q_n}(\mathbf{H}_p^r)_q$ differ by a factor of the order $n^{(d-1)c(p,q)}$.

In the case $p = q = 1$ the classes \mathbf{W} and \mathbf{H} are alike in the sense of best approximation by the hyperbolic cross polynomials (see Theorems 4.5 and 4.6 above):

$$E_{Q_n}(\mathbf{W}_1^r)_1 \asymp E_{Q_n}(\mathbf{H}_1^r)_1 \asymp 2^{-rn} n^{d-1}.$$

It turns out that approximation in the uniform metric differs essentially from approximation in the L_q -metric, $1 < q < \infty$, not only in the methods of proof, but also in that the results are fundamentally different. For example, in approximation in the L_q -metric, $1 < q < \infty$, the partial Fourier sums $S_{Q_n}(f)$ give the order of the best approximation $E_{Q_n}(f)_q$ and thus, if we are not interested in the dependence of $E_{Q_n}(f)_q$ on q , then we can confine ourselves to the study of $S_{Q_n}(f)$.

In the univariate case and the uniform metric the partial sums of the Fourier series give good approximation for the functions in the classes W_p^r and H_p^r , $1 < p < \infty$:

$$E_n(F)_\infty \asymp \sup_{f \in F} \|f - S_n(f)\|_\infty,$$

where F denotes either W_p^r or H_p^r .

In the case of the classes \mathbf{W}_p^r and \mathbf{H}_p^r , $1 < p < \infty$, not only the Fourier sums do not give the orders of the least upper bounds of the best approximations in the L_∞ -norm, but also no linear method gives the orders of the least upper bounds of the best approximations with respect to the classes \mathbf{W}_p^r or \mathbf{H}_p^r , $d \geq 2$, $1 < p < \infty$ (see [340], Chapter 2, Section 5). In other words, the operator of the best approximation in the uniform metric by polynomials in $\mathcal{T}(Q_n)$ cannot be replaced by any linear operator without detriment to the order of approximation on the classes \mathbf{W}_p^r and \mathbf{H}_p^r , $1 < p < \infty$.

Let us continue with results on the Besov class $\mathbf{B}_{p,\theta}^r$. We will see how the third parameter θ in this class is reflected on the asymptotic order of $E_{Q_n}(\mathbf{B}_{p,\theta}^r)_q$.

Theorem 4.9. *Let $1 < p, q < \infty$, $1 \leq \theta < \infty$, $r > (1/p - 1/q)_+$. Then we have*

$$E_{Q_n}(\mathbf{B}_{p,\theta}^r)_q \asymp \begin{cases} 2^{-(r-1/p+1/q)n} n^{(1/q-1/\theta)+(d-1)}, & p < q; \\ 2^{-rn}, & q \leq p, \theta \leq \min\{2, p\}; \\ 2^{-rn} n^{(1/2-1/\theta)(d-1)}, & q \leq p, p \geq 2, \theta > 2; \\ 2^{-rn} n^{(1/p-1/\theta)(d-1)}, & p = q < 2, \theta > p. \end{cases}$$

Theorem 4.9 was proved in [87]. The upper bounds are realized by the approximation by the operator S_{Q_n} . Although in the case $q \leq p$, $\theta < \min\{2, p\}$, $\mathbf{B}_{p,\theta}^r \subsetneq \mathbf{W}_p^r$, we still have $E_{Q_n}(\mathbf{B}_{p,\theta}^r)_q \asymp E_{Q_n}(\mathbf{W}_p^r)_q$. While in the case $q \leq p$, $\theta > \min\{2, p\}$ where $\mathbf{B}_{p,\theta}^r \not\subset \mathbf{W}_p^r$ and the approximation properties of $\mathbf{B}_{p,\theta}^r$ are closer to those of \mathbf{H}_p^r , the asymptotic order of $E_{Q_n}(\mathbf{B}_{p,\theta}^r)_q$ has the additional logarithm term $n^{(1/\min\{2,p\}-1/\theta)(d-1)}$.

Theorem 4.10. *Let $1 \leq p < \infty$, $1 \leq \theta < \infty$, $r > 0$. Then we have*

$$E_{Q_n}(\mathbf{B}_{p,\theta}^r)_1 \asymp \begin{cases} 2^{-rn} n^{(1/p-1/\theta)+(d-1)}, & p \leq 2, \\ 2^{-rn}, & p > 2, \theta \leq 2. \end{cases}$$

Theorem 4.10 was proved in the case $p = 1$ in [273] and in the case $1 < p \leq 2$ in [277]. In the case $p > 2$ and $\theta \leq 2$ the upper bound follows from the embedding of the $\mathbf{B}_{p,\theta}^r$ into \mathbf{W}_2^r and Theorem 4.4. The lower bound in this case is trivial. Similarly to Theorem 4.9, the upper bounds in this theorem are realized by the approximation by the operator S_{Q_n} . The lower bounds are proved by the construction of a “fooling” test function.

Theorem 4.11. *Let $1 \leq p \leq 2$, $1 \leq \theta \leq 2$, $r > 1/p$. Then we have*

$$E_{Q_n}(\mathbf{B}_{p,\theta}^r)_\infty \asymp 2^{-(r-1/p)n}. \quad (4.3)$$

Theorem 4.11 was proved in [277]. The upper bound follows from the embedding of the $\mathbf{B}_{p,\theta}^r$ into $\mathbf{W}_2^{r-1/p+1/2}$ and Theorem 4.5. The lower bound follows from relation (4.3) for the univariate case ($d = 1$).

We note that in some cases the upper bounds are trivial. For instance, in the case $\theta = 1$, using the Nikol'skii inequalities, we obtain for $1 \leq p \leq q \leq \infty$, $r > 1/p - 1/q$, that

$$\begin{aligned} E_{Q_n}(f)_q &\leq \sum_{|\mathbf{s}|_1 \geq n} \|A_{\mathbf{s}}(f)\|_q \lesssim \sum_{|\mathbf{s}|_1 \geq n} 2^{|\mathbf{s}|_1(1/p-1/q)} \|A_{\mathbf{s}}(f)\|_p \\ &\leq 2^{(-r+1/p-1/q)n} \sum_{|\mathbf{s}|_1 \geq n} 2^{r|\mathbf{s}|_1} \|A_{\mathbf{s}}(f)\|_p \ll 2^{(-r+1/p-1/q)n} \|f\|_{\mathbf{B}_{p,1}^r}. \end{aligned}$$

The corresponding lower bounds follow from the univariate case. Thus, we obtain for $1 \leq p \leq q \leq \infty$, $r > 1/p - 1/q$ (see [87] for $q < \infty$ and [278] for $q = \infty$)

$$E_{Q_n}(\mathbf{B}_{p,1}^r)_q \asymp 2^{(-r+1/p-1/q)n}.$$

In other cases the lower bounds follow from the corresponding examples, used for the \mathbf{H} classes, and a simple inequality for $f \in \mathcal{T}(\Delta Q_n)$

$$\|f\|_{\mathbf{B}_{p,\theta}^r} \lesssim n^{(d-1)/\theta} \|f\|_{\mathbf{H}_p^r}. \quad (4.4)$$

For instance, in this way we obtain (see [278])

$$E_{Q_n}(\mathbf{B}_{\infty,\theta}^r)_q \asymp 2^{-rn} n^{(\frac{1}{2}-\frac{1}{\theta})_+^{(d-1)}}$$

for $1 < q < \infty$, $r > 0$, $1 \leq \theta < \infty$;

$$E_{Q_n}(\mathbf{B}_{p,\theta}^r)_1 \asymp 2^{-rn} n^{(\frac{1}{2}-\frac{1}{\theta})}$$

if $d = 2$ and $2 < \theta < \infty$, $2 < p \leq \infty$, $r > 0$; and (see [279])

$$E_{Q_n}(\mathbf{B}_{\infty,\theta}^r)_\infty \asymp 2^{-rn} n^{1-\frac{1}{\theta}}$$

if $d = 2$ and $1 \leq \theta < \infty$, $r > 0$.

Here is a result from [276].

Theorem 4.12. *Let $1 < q < \infty$, $r > 1 - \frac{1}{q}$, $1 \leq \theta < \infty$. Then we have*

$$E_{Q_n}(\mathbf{B}_{1,\theta}^r)_q \asymp 2^{-\left(r-1+\frac{1}{q}\right)n} n^{\left(\frac{1}{q}-\frac{1}{\theta}\right)_+^{(d-1)}}.$$

We note that the upper bounds in Theorem 4.12 are derived from Theorem 2.11.

4.3 The Kolmogorov widths

We begin with results on \mathbf{W}_p^r classes. Denote

$$r(p, q) := \begin{cases} (1/p - 1/q)_+ & \text{for } 1 \leq p \leq q \leq 2; \\ \max\{1/2, 1/p\} & \text{otherwise.} \end{cases} \quad 1 \leq q \leq p \leq \infty,$$

Theorem 4.13. *Let $r(p, q)$ be as above. Then for $1 < p, q < \infty$, $r > r(p, q)$*

$$d_m(\mathbf{W}_p^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r - \left(\frac{1}{p} - \max\{1/2, 1/q\} \right)_+}.$$

As we already mentioned in the Introduction in the case $p = q = 2$ Theorem 4.13 follows from a general result by Babenko [6] and in the case $1 < p = q < \infty$ it was proved by Mityagin [233]. Note that in [233] only the case of natural r was considered. The result was extended to real r in [131]. In the case $1 < p < q \leq 2$ the theorem was obtained in [324] and [331]. In the case $1 < p < q < \infty$, $2 \leq q < \infty$ it was obtained in [325] and [334]. In the case $1 < q < p < \infty$ Theorem 4.13 was proved in [135]. In all cases $1 < p \leq q \leq 2$ and $1 < q \leq p < \infty$, included in Theorem 4.13, the upper estimates follow from approximation by the hyperbolic cross polynomials from $\mathcal{T}(Q_n)$ with $m \asymp |Q_n|$, $|Q_n| \leq m$.

In the case $2 \leq p \leq q < \infty$ we encounter an interesting and important phenomenon in Theorem 4.13. The main rate of convergence is r and the rate does neither depend on p nor on q . In the univariate case this effect has been first observed by Kashin in his seminal paper [192]. It makes use of the Maiorov discretization technique [225] where the problem of n -widths for function classes is reduced to the study of n -widths in finite dimensional normed spaces, see Theorem 4.16 below. In the multivariate case (suppose $2 \leq p \leq q < \infty$) the proof of Theorem 4.13 (and Theorem 4.17 below) is based on the following result.

Theorem 4.14. *One has the estimate*

$$d_m(\mathcal{T}(\mathbf{N}, d)_2, L_\infty) \lesssim (\vartheta(\mathbf{N})/m)^{1/2} \ln(e\vartheta(\mathbf{N})/m)^{1/2}.$$

We illustrate on the example of estimating from above the $d_m(\mathbf{W}_2^r, L_q)$, $2 \leq q < \infty$, how Theorem 4.14 is applied. First of all, we derive the following lemma from Theorem 4.14.

Lemma 4.15. *Let $2 \leq q < \infty$. We have for $\Delta Q_n := Q_n \setminus Q_{n-1}$*

$$d_m(\mathcal{T}(\Delta Q_n)_2, L_q) \lesssim (|\Delta Q_n|/m)^{1/2} (\ln(e|\Delta Q_n|/m))^{1/2}.$$

Proof. By Corollaries 1 and 2 of the Littlewood-Paley theorem it is easy to see that

$$d_m(\mathcal{T}(\Delta Q_n)_2, L_q) \lesssim \max_{\mathbf{s} \in \theta_n} d_{[m/|\theta_n|]}(\mathcal{T}(\rho(\mathbf{s}))_2, L_q),$$

where $\theta_n := \{\mathbf{s} : |\mathbf{s}|_1 = n\}$. Applying Theorem 4.14 we obtain Lemma 4.15. ■

Second, let $r > 1/2$ and $\kappa > 0$ be such that $r > 1/2 + \kappa$. For $n \in \mathbb{N}$ set

$$m_l := \lfloor |\Delta Q_n| 2^{\kappa(n-l)} \rfloor, \quad l = n+1, \dots;$$

$$m := |Q_n| + \sum_{l>n} m_l \leq C(\kappa, d) 2^n n^{d-1}.$$

Then

$$d_m(\mathbf{W}_2^r, L_q) \leq \sum_{l>n} d_{m_l}(\mathcal{T}(\Delta Q_l)_2, L_q) 2^{-rl} \lesssim 2^{-rn} \asymp m^{-r} (\log m)^{(d-1)r}.$$

Theorem 4.14 is a corollary of the following fundamental result of Kashin [192], Gluskin [144] and Garnaev, Gluskin [142]. See also [52], [256] and the recent papers [398], [124] for dual versions of the result.

Theorem 4.16. *For any natural numbers d, m , $m < d$ we have*

$$d_m(B_2^d, \ell_\infty^d) \leq C m^{-1/2} (\ln(ed/m))^{1/2}.$$

Theorem 4.14 is derived from Theorem 4.16 by discretization using the multivariate version of the Marcinkiewicz Theorem 2.17 from Section 2. We note here that it would be natural to try to apply the discretization technique to the hyperbolic cross polynomials in $\mathcal{T}(Q_n)$. However, it follows from the discussion in Subsection 2.5 of Section 2 that this way does not work. This brings technical difficulties in the analysis of the case $2 < q < \infty$.

We note that the proof of Theorem 4.16 both in [192] and in [142] is probabilistic. The authors prove existence of matrices with special properties. This makes Theorem 4.16 and all results obtained with its help non-constructive. It turns out that matrices similar to those constructed in [192] and [142] are very important in sparse approximation, namely, in compressed sensing [125], [368]. Such matrices are called the matrices with the Restricted Isometry Property (RIP) or simply the RIP matrices. The reader can find a discussion of such matrices and an introduction to compressed sensing in [125, Chapt. 6 & 9], [368, Chapt. 5], [124]. We point out that the problem of deterministic construction of good (near optimal) RIP matrices is a fundamental open problem of the compressed sensing theory, see [125, p. 170].

The lower bounds in the case $1 < p < q \leq 2$ in Theorem 4.13 required a new technique. A new element of the technique is that in the multivariate case a subspace, from which we choose a “bad” function, depends on the system $\varphi_1, \dots, \varphi_m$ of functions used for approximation in the definition of the Kolmogorov width. In the univariate case we can always take a “bad” function from $\mathcal{T}(n)$ with appropriate $n \asymp m$.

Let us now proceed with some limiting cases.

Theorem 4.17. *Let $r > 1$ and $2 \leq q < \infty$. Then*

$$d_m(\mathbf{W}_1^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/2} (\log m)^{(d-1)/2}.$$

This theorem was obtained in [325] and [334].

Theorem 4.18. *For all $1 \leq q < \infty$ and $r > 0$ we have*

$$d_m(\mathbf{W}_\infty^r, L_q) \asymp m^{-r} (\log m)^{r(d-1)}.$$

The upper bounds in Theorem 4.18 follow from the case $1 < p = q < \infty$. The corresponding lower bounds in Theorem 4.18 were proved in [195] in the case $1 < q < \infty$ and in [196] in the case $q = 1$. These proofs are based on finite dimensional geometry results on volume estimates (see Subsection 2.5 of Section 2). As a corollary of the lower bound in Theorem 4.18 and upper bounds from Theorem 4.13 we obtain the following result.

Theorem 4.19. *For all $1 < p \leq \infty$ and $r > 0$ we have*

$$d_m(\mathbf{W}_p^r, L_1) \asymp m^{-r} (\log m)^{r(d-1)}.$$

The first result on the right order of $d_m(\mathbf{W}_p^r, L_\infty)$ was obtained in [356] in the case $d = 2$.

Theorem 4.20. *Let $d = 2$ and $2 \leq p \leq \infty$, $r > 1/2$. Then*

$$d_m(\mathbf{W}_p^r, L_\infty) \asymp m^{-r} (\log m)^{r+1/2}.$$

The most difficult part of Theorem 4.20 is the lower bounds. The proof of the lower bounds is based on the Small Ball Inequality (see (2.27)) from Section 2). In the case $2 \leq p < \infty$ Theorem 4.20 is proved in [356] and in the case $p = \infty$ in [357]. The region

$$R_1 := \{(1/p, 1/q) : 0 < 1/p \leq 1/q < 1 \text{ or } 1/2 \leq 1/q \leq 1/p < 1\}$$

is covered by Theorem 4.13. For this region the upper bounds follow from Theorem 4.4, which means that in this case the subspaces $\mathcal{T}(Q_n)$ of the hyperbolic cross polynomials are optimal in the sense of order. Theorem 4.13 shows that for the region $R_2 := (0, 1)^2 \setminus R_1$ the subspaces $\mathcal{T}(Q_n)$ are not optimal in the sense of order. Theorem 4.13 gives the orders of the $d_m(\mathbf{W}_p^r, L_q)$ for all $(1/p, 1/q)$ from the open square $(0, 1)^2$ under some restrictions on r . The situation on the boundary of this square is more difficult. Theorem 4.17 covers the segment $S_1 := \{(1, 1/q) : 0 < 1/q \leq 1/2\}$. The segment $S_2 := \{(0, 1/q) : 0 < 1/q \leq 1\}$ is covered by Theorem 4.18. The segment $S_3 := \{(1/p, 1) : 0 \leq 1/p < 1\}$ is covered by Theorem 4.19. Finally, the segment $S_4 := \{(1/p, 0) : 0 \leq 1/p \leq 1/2\}$ in the case $d = 2$ is covered by Theorem 4.20. In all other cases the right order of the $d_m(\mathbf{W}_p^r, L_q)$ is not known.

Let us discuss an extension of Theorem 4.20 to the case $d \geq 3$. The following upper bounds are known (see [31], the book [385], and [116] for the special case $r = 1$)

$$d_m(\mathbf{W}_p^r, L_\infty) \lesssim m^{-r}(\log m)^{(d-1)r+1/2}, \quad 2 \leq p \leq \infty, \quad r > 1/2.$$

Recent results on the Small Ball Inequality for the Haar system (see [36], [37]) allow us to improve a trivial lower bound to the following one for $r = 1$ and all $p < \infty, d \geq 3$:

$$d_m(\mathbf{W}_p^1, L_\infty) \gtrsim m^{-1}(\log m)^{d-1+\delta(d)}, \quad \delta(d) > 0.$$

Theorem 4.20 and the above upper and lower bounds support the following

Conjecture 4.21. *Let $d \geq 3, 2 \leq p \leq \infty, r > 1/2$. Then*

$$d_m(\mathbf{W}_p^r, L_\infty) \asymp m^{-r}(\log m)^{(d-1)r+1/2}.$$

We summarize the above results on the $d_m(\mathbf{W}_p^r, L_q)$ in the following figure.

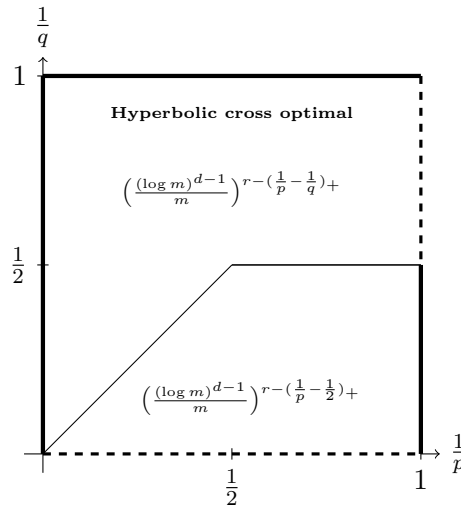


Figure 4: The asymptotical order of $d_m(\mathbf{W}_p^r, L_q)$

A straight line on the boundary indicates that in the respective parameter region the correct order is known. In the dashed line region we do not know the correct order.

We now proceed to classes \mathbf{H}_p^r . The first result on the right order of the Kolmogorov width for \mathbf{H}_p^r classes was obtained in [324]. The proper lower bound for $d_m(\mathbf{H}_\infty^r, L_2)$ was proved in [324].

Theorem 4.22. *Let $r(p, q)$ be the same as in Theorem 4.13. Then for $1 < p \leq q < \infty$, $r > r(p, q)$*

$$d_m(\mathbf{H}_p^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r - \left(\frac{1}{p} - \max\{1/2, 1/q\} \right)_+} (\log m)^{(d-1) \max\{1/2, 1/q\}}.$$

In the case $1 \leq p < 2$, $q = 2$ Theorem 4.22 was obtained in [331]. In the case $1 < p < q < \infty$, $q \geq 2$ – in [325] and [334]. In the case $1 < p \leq q < 2$ – in [138].

Theorem 4.23. *Let $1 < q < p \leq \infty$, $p \geq 2$ and $r > 0$. Then*

$$d_m(\mathbf{H}_p^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(d-1)/2}.$$

In the case $2 \leq q \leq p \leq \infty$, $q < \infty$, Theorem 4.23 was obtained in [324]. In the case $1 < q < 2 \leq p < \infty$ Theorem 4.23 was obtained in [135], [134] and in [86]. In the case $p = \infty$ and $1 < q < \infty$ – in [342] and [344].

Theorem 4.24. *Let $2 \leq q < \infty$, $r > 1$. Then*

$$d_m(\mathbf{H}_1^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/2} (\log m)^{(d-1)/2}.$$

This theorem is from [325] and [334].

Theorem 4.25. *Let $2 \leq p \leq \infty$, $r > 0$. Then*

$$d_m(\mathbf{H}_p^r, L_1) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(d-1)/2}.$$

The upper bounds in Theorem 4.25 follow from the upper bounds for $E_{Q_n}(\mathbf{H}_2^r)_2$ from Theorem 4.6. It was observed in [31] that the lower bounds follow from the corresponding lower bounds for the entropy numbers $\epsilon_k(\mathbf{H}_\infty^r, L_1)$ from [342] and [344].

Theorem 4.26. *Let $d = 2$, $2 \leq p \leq \infty$ and $r > 1/2$ when $p < \infty$, and $r > 0$ when $p = \infty$. Then*

$$d_m(\mathbf{H}_p^r, L_\infty) \asymp m^{-r} (\log m)^{r+1}.$$

The most difficult part of Theorem 4.26 is the lower bounds for $p = \infty$. The proof of the lower bounds is based on the Small Ball Inequality (see (2.27) from Section 2). Theorem 4.26 is proved in [356] with the assumption $r > 0$ for the lower bounds in case $p = \infty$ and with the assumption $r > 1/2$ for the upper bounds. We note that in case $p = \infty$ the matching upper bounds follow from Theorem 4.8 under assumption $r > 0$.

In case $d \geq 3$ the right order of the $d_m(\mathbf{H}_p^r, L_\infty)$, $2 \leq p \leq \infty$, is not known. The following upper bounds are known (see [31, 32])

$$d_m(\mathbf{H}_p^r, L_\infty) \lesssim m^{-r} (\log m)^{(d-1)(r+1/2)+1/2}, \quad 2 \leq p \leq \infty, \quad r > 1/2.$$

Theorem 4.26 and the above upper bounds support the following conjecture.

Conjecture 4.27. *Let $d \geq 3$, $2 \leq p \leq \infty$, $r > 1/2$. Then*

$$d_m(\mathbf{H}_p^r, L_\infty) \asymp m^{-r} (\log m)^{(d-1)(r+1/2)+1/2}.$$

We summarize the above results on the $d_m(\mathbf{H}_p^r, L_q)$ in the following picture.

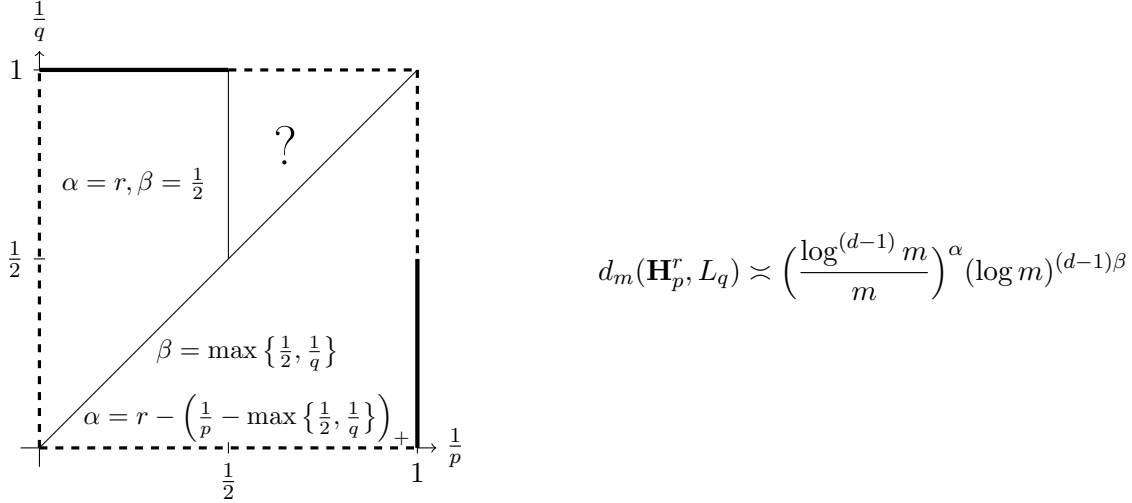


Figure 5: The asymptotical order of $d_m(\mathbf{H}_p^r, L_q)$

The region $P_1 := \{(1/p, 1/q) : 0 \leq 1/p < 1/q < 1, 1/p \leq 1/2\}$ is covered by Theorem 4.23. For this region the upper bounds follow from Theorem 4.6. Therefore, in this case the subspaces $\mathcal{T}(Q_n)$ are optimal in the sense of order. In the case of region $P_2 := \{(1/p, 1/q) : 1/2 < 1/p < 1/q \leq 1\}$ the right order of $d_m(\mathbf{H}_p^r, L_q)$ is not known. The region $P_3 := \{(1/p, 1/q) : 0 < 1/q \leq 1/p < 1\}$ is covered by Theorem 4.22. The boundary segment $A_1 := \{(1, 1/q) : 0 < 1/q \leq 1/2\}$ is covered by Theorem 4.24, the segment $A_2 := \{(1/p, 1) : 0 \leq 1/p \leq 1/2\}$ is covered by Theorem 4.25, and, finally, the segment $A_3 := \{(1/p, 0) : 0 \leq 1/p \leq 1/2\}$ in the case $d = 2$ is covered by Theorem 4.26. In all other cases the right orders of $d_m(\mathbf{H}_p^r, L_q)$ are not known.

Let us continue with results on $d_m(\mathbf{B}_{p,\theta}^r, L_q)$. As for $E_{Q_n}(\mathbf{B}_{p,\theta}^r)_q$, the third parameter θ is reflected on the asymptotic order of $d_m(\mathbf{B}_{p,\theta}^r, L_q)$.

Theorem 4.28. *Let $r > 0$, $1 < p \leq q < \infty$, $1 \leq \theta < \infty$, $\beta := \frac{1/p-1/q}{1-2/q}$. Then we have*

$$d_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-[1/p-\max\{1/2, 1/q\}]_+} (\log m)^{(d-1)\delta},$$

where

$$\delta := \begin{cases} (1/2 - 1/\theta)_+, & \text{if } 2 \leq p < q, \ r > \beta, \\ [\max\{1/2, 1/q\} - 1/\theta]_+, & \text{if } p \leq q \leq 2, \ r > 1/p - 1/q, \text{ or} \\ & p \leq 2 < q, \ r > 1/p. \end{cases}$$

In Theorem 4.28, the cases $2 \leq p < q$ and $p \leq 2 < q$ were proved in [141], the upper bound of the case $p \leq q \leq 2$ was proved in [87, 88], the lower bound of the case $p \leq q \leq 2$ was proved in [269] for $\theta \geq 1$.

Theorem 4.29. *Let $r > 0$, $1 \leq \theta < \infty$. Let $2 \leq q \leq p < \infty$ or $1 < q < 2 \leq p < \infty$ and $\theta \geq 2$. Then we have*

$$d_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(d-1)(1/2-1/\theta)_+}.$$

In Theorem 4.29, the upper bounds were proved in [87, 88], the lower bounds were proved in [141].

Let us give a brief comment on the proofs of Theorems 4.28 and 4.29 for which $1 < p, q < \infty$. For details, the reader can see [87, 88, 141, 269]. We first treat the upper bounds. The cases $p \leq q \leq 2$, $2 \leq q \leq p < \infty$ and $1 < q < 2 \leq p < \infty$ are derived from the linear approximation by the operators S_{Q_n} , the case $p \leq 2 < q$ can be reduced by the embedding $\mathbf{B}_{p,\theta}^r \hookrightarrow \mathbf{B}_{2,\theta}^{r-1/p+1/2}$ (see Lemma 3.12) to the case $2 \leq p < q$. By using Corollary 11.8 the case $2 \leq p < q$ can be reduced to the upper bounds of Kolmogorov widths of finite-dimensional sets which can be estimated by the following Kashin-Gluskin's Lemma 4.30, see [192], [144] below. Below, in Lemmas 4.30-4.34, Theorem 4.16 and Lemma 4.51 we will state results on n -widths and linear widths of finite dimensional sets. For a rather complete survey on these results see e.g. Vybíral [398].

Lemma 4.30. *Let $2 \leq p < q < \infty$ and $1/p + 1/q \geq 1$ and $\beta := \frac{1/p-1/q}{1-2/q}$. Then we have for $n > m$,*

$$d_m(B_p^n, \ell_q^n) \asymp \min\{1, n^{2\beta/q} m^{-\beta}\}.$$

Notice also that the restriction on the smoothness $r > \beta$ in the case $2 \leq p < q$ is important in the proof of the upper bound, for details see [144].

We next consider the lower bounds in Theorem 4.29. The cases $p \leq q \leq 2$, $2 \leq q \leq p < \infty$ and $1 < q < 2 \leq p < \infty$ can be reduced to the Kolmogorov widths of finite-dimensional sets which can be upper estimated for the cases $2 \leq q \leq p < \infty$ and $1 < q < 2 \leq p < \infty$ by Pietsch-Stesin's lemma below. The cases $2 \leq p < q$, $p \leq 2 < q$ can be reduced by the norm inequality $\|\cdot\|_q \geq \|\cdot\|_2$ to the cases $p \leq q \leq 2$, $p \geq q \geq 2$ with $q = 2$.

Lemma 4.31. *Let $1 \leq q \leq p \leq \infty$. Then we have for $n > m$,*

$$d_m(B_p^n, \ell_q^n) \asymp (n-m)^{1/q-1/p}.$$

For the case $p \leq q \leq 2$, the upper bounds of Kolmogorov widths of finite-dimensional sets are estimated by Kolmogorov-Petrov-Smirnov-Mal'tsev-Stechkin's equation $d_m(B_1^n, \ell_2^n) = (1 - m/n)^{-1/2}$ (see, e.g., [141] for references).

Another important tool is given by Galeev's lemma [141] below. Let for $1 \leq p, q \leq \infty$ the finite dimensional normed space $\ell_{p,q}^{n,s}$ of vectors $\mathbf{x} \in \mathbb{R}^{sn}$ be equipped with the norm

$$\|x\|_{\ell_{p,q}^{n,s}} := \left(\sum_{j=1}^s \left(\sum_{(j-1)n < k \leq jn} |x_k|^p \right)^{q/p} \right)^{1/q}.$$

$B_{p,q}^{n,s}$ denotes its unit ball.

Lemma 4.32. *Let $1 < q < \infty$. Then we have for $m \leq ns/2$ that*

$$d_m(B_{1,\infty}^{n,s}, \ell_{2,q}^{n,s}) \asymp s^{1/q}.$$

Izaak [182, 183] extended the above result to $q = 1$ and obtained the following.

Lemma 4.33. *For $m \leq ns/2$ it holds*

$$s \frac{\sqrt{\log \log s}}{\log s} \lesssim d_m(B_{1,\infty}^{n,s}, \ell_{2,1}^{n,s}) \lesssim s.$$

The case $q = 1$ is particularly important for proving some of the sharp lower bounds for the linear widths of the Hölder-Nikolskii classes in Theorem 4.46 below. See the proof of the main theorem in [140]. Note that there is a very tiny gap in the result by Izaak which has been closed only recently by Malykhin and Ryutin [229] with a new method based on a probabilistic argument. In fact, we have

Lemma 4.34. *For $m \leq ns/2$ it holds*

$$d_m(B_{1,\infty}^{n,s}, \ell_{2,1}^{n,s}) \asymp s.$$

Theorem 4.35. *Let $1 < q \leq 2$, $r > 1 - 1/q$, $1 \leq \theta \leq q$ or $2 \leq q < \infty$, $r > 1$, $1 \leq \theta < \infty$. Then we have*

$$d_m(\mathbf{B}_{1,\theta}^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r - [1 - \max\{1/2, 1/q\}]_+} (\log m)^{(d-1)\delta},$$

where

$$\delta := \begin{cases} (1/2 - 1/\theta)_+, & \text{if } 2 \leq q < \infty, \\ 0, & \text{if } 1 < q \leq 2. \end{cases}$$

Theorem 4.35 is proved in [274].

Theorem 4.36. *Let $r > 0$. Then we have*

$$d_m(\mathbf{B}_{\infty,1}^r, L_\infty) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r.$$

Theorem 4.36 was proved in [275]. Notice that in both Theorems 4.35 and 4.36 the upper bounds are derived from the linear approximation by the operators V_{Q_n} . We note that the following relation

$$d_m(\mathbf{B}_{\infty,1}^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r, \quad 2 \leq q \leq \infty$$

holds. The upper bounds are derived from the linear approximation by the operators V_{Q_n} . The lower bounds follow from Example 1 of Chapter 3 in [352].

4.4 Orthowidths

We begin with the classes \mathbf{W}_p^r . For numbers m and r and parameters $1 \leq p, q \leq \infty$ we define the functions

$$w(m, r, p, q) := (m^{-1}(\log m)^{d-1})^{r - (1/p - 1/q)_+} (\log m)^{(d-1)\xi(p,q)},$$

where $(a)_+ := \max\{a, 0\}$ and

$$\xi(p, q) = \begin{cases} 0 & \text{for } 1 < p \leq q < \infty, & 1 \leq q < p \leq \infty, \\ 1 - 1/p & \text{for } 1 < p \leq \infty, & q = \infty, \\ 1/q & \text{for } p = 1, & 1 \leq q \leq \infty. \end{cases}$$

Theorem 4.37. *Suppose that $r > (1/p - 1/q)_+$ and $(p, q) \neq (1, 1), (\infty, \infty), (1, \infty)$. Then*

$$\varphi_m(\mathbf{W}_p^r, L_q) \asymp w(m, r, p, q),$$

and orthogonal projections S_{Q_n} onto the subspaces $\mathcal{T}(Q_n)$ of trigonometric polynomials with harmonics in the step hyperbolic crosses with n such that $|Q_n| \leq m$, $|Q_n| \asymp m$ give the order of the quantities $\varphi_m(\mathbf{W}_p^r, L_q)$.

Theorem 4.37 in the case $1 < p \leq q < \infty$ was obtained in [325] and [334] and in the general case in [342] and [344].

The following theorem from [5] covers the case $(p, q) = (1, 1)$ or (∞, ∞) .

Theorem 4.38. *The following order estimates hold for $p = 1$ and $p = \infty$*

$$\varphi_m(\mathbf{W}_p^r, L_p) \asymp m^{-r}(\log m)^{(d-1)(r+1)}, \quad r > 0.$$

The only case, which is not completely studied, is the case $p = 1$ and $q = \infty$. We have here the following partial result, which shows an interesting behavior of $\varphi_m(\mathbf{W}_1^r, L_\infty)$ in this case.

Theorem 4.39. *For odd $r > 1$ and arbitrary d we have*

$$\varphi_m(\mathbf{W}_1^r, L_\infty) \asymp (m^{-1}(\log m)^{d-1})^{r-1},$$

and for even r and $d = 2$ we have

$$\varphi_m(\mathbf{W}_1^r, L_\infty) \asymp m^{1-r}(\log m)^r.$$

We now proceed to classes \mathbf{H}_p^r . For numbers m and r and parameters $1 \leq p, q \leq \infty$ we define the functions

$$h(m, r, p, q) = (m^{-1}(\log m)^{d-1})^{r-(1/p-1/q)_+} (\log m)^{(d-1)\eta(p,q)},$$

where

$$\eta(p, q) = \begin{cases} 1/q & \text{for } 1 < p < q < \infty; & p = 1, & 1 \leq q < \infty; \\ 1 & \text{for } 1 \leq p \leq \infty, & q = \infty; \\ 1/2 & \text{for } 1 \leq q \leq p \leq \infty, & p \geq 2, & q < \infty; \\ 1/q & \text{for } 1 \leq q \leq p \leq 2. \end{cases}$$

Theorem 4.40. *Suppose that $r > (1/p - 1/q)_+$ and $(p, q) \neq (1, 1), (\infty, \infty)$. Then*

$$\varphi_m(\mathbf{H}_p^r, L_q) \asymp h(m, r, p, q)$$

and subspaces, optimal in the sense of order, are given by $\mathcal{T}(Q_n)$ with appropriate n (as in Theorem 4.37).

Theorem 4.40 in the case $1 \leq p < q < \infty$ and $1 < p = q \leq 2$ was obtained in [325] and [334]. In the case $1 < q < p < 2$ it was proved in [136]. In the general form it was obtained in [342] and [344].

The following theorem from [5] covers the case $(p, q) = (1, 1)$ or (∞, ∞) .

Theorem 4.41. *The following order estimates hold for $p = 1$ and $p = \infty$*

$$\varphi_m(\mathbf{H}_p^r, L_p) \asymp m^{-r}(\log m)^{(d-1)(r+1)}, \quad r > 0.$$

In the cases $p = q = 1$ and $p = q = \infty$ the operators S_{Q_n} do not provide optimal in the sense of order approximation for classes \mathbf{W} and \mathbf{H} . In these cases other orthonormal system – the wavelet type system of orthogonal trigonometric polynomials – was used (see [5]). The proofs of upper bounds in Theorems 4.38 and 4.41 are based on the approach discussed in the beginning of Subsection 4.2 (see Proposition 4.1).

The main point of the proof of Theorems 4.37 and 4.40 is in lower bounds. The lower bound proofs are based on special examples. Some of these examples are simple, like, $e^{i(\mathbf{k}^0, \mathbf{x})}$ with \mathbf{k}^0 determined by the system u_1, u_2, \dots, u_m from the definition of the orthowidth φ_m . Other examples are more involved. For instance, a function $g(\mathbf{x} + \mathbf{y}^*)$ of the form

$$g(\mathbf{x}) := \sum_{\mathbf{s} \in \theta_n^1} e^{i(\mathbf{k}^{\mathbf{s}}, \mathbf{x})},$$

with \mathbf{y}^* , $\{\mathbf{k}^{\mathbf{s}}\}_{\mathbf{s} \in \theta_n^1}$, $\theta_n^1 \subset \theta_n := \{\mathbf{s} : |\mathbf{s}|_1 = n\}$ are determined by the system u_1, u_2, \dots, u_m . The reader can find a detailed discussion of these examples in [344] and [352].

We note that the case $1 \leq q < p < 2$ turns out to be difficult for classes \mathbf{H}_p^r . Even the corresponding result for $E_{Q_n}(\mathbf{H}_p^r)_q$ was difficult and required a new technique. The right order of the Kolmogorov widths in this case is still unknown.

We now proceed to classes $\mathbf{B}_{p,\theta}^r$. For numbers m and r and parameters $1 < p, q < \infty$, $1 \leq \theta < \infty$, we define the functions

$$g(m, r, p, q, \theta) = \left(\frac{(\log m)^{d-1}}{m} \right)^{r-(1/p-1/q)_+} (\log m)^{(d-1)\delta(p,q,\theta)},$$

where

$$\delta(p, q, \theta) = \begin{cases} (1/q - 1/\theta)_+ & \text{if } p < q; \\ 1 & \text{if } q \leq p, \theta \leq \min\{p, 2\}; \\ 1/2 - 1/\theta & \text{if } q \leq p, p \geq 2, \theta \geq 2; \\ (1/p - 1/\theta)_+ & \text{if } q \leq p \leq 2. \end{cases}$$

Theorem 4.42. *Let $1 < p, q < \infty$, $1 \leq \theta < \infty$, $r > (1/p - 1/q)_+$. Then*

$$\varphi_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp g(m, r, p, q, \theta)$$

and subspaces, optimal in the sense of order, are given by $\mathcal{T}(Q_n)$ with appropriate n (as in Theorem 4.37).

Theorem 4.42 was obtained in [88] excepting the cases $q < p \leq 2$ and $q = p \leq 2, q > \theta$. The upper bounds of the last two cases were proved in [87], the lower bounds follow from a result on nuclear widths in [139] (the nuclear width which is smaller than orthowidth is defined in a way similar to the definition of orthowidth by replacing orthogonal projectors by more general nuclear operators, for details see [139]).

Theorem 4.43. *Let $1 < p \leq \infty$, $1 \leq \theta < \infty$, $r > 0$. Then*

$$\varphi_m(\mathbf{B}_{p,\theta}^r, L_1) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(d-1)(1/p-1/\theta)_+}$$

and subspaces, optimal in the sense of order, are given by $\mathcal{T}(Q_n)$ with appropriate n (as in Theorem 4.37).

Theorem 4.43 was proved in [277].

The lower bounds (the main part) in the following result (see [279]) follow from (4.4) and the corresponding example for the class \mathbf{H}_p^r .

Theorem 4.44. *Let $1 \leq p < \infty$, $r > \frac{1}{p}$, $1 \leq \theta < \infty$. Then*

$$\varphi_m(\mathbf{B}_{p,\theta}^r, L_\infty) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-\frac{1}{p}} (\log m)^{(d-1)(1-\frac{1}{\theta})},$$

and subspaces, optimal in the sense of order, are given by $\mathcal{T}(Q_n)$ with appropriate n (as in Theorem 4.32).

The right orders of the $\varphi_m(\mathbf{B}_{p,\theta}^r, L_q)$ in the cases $p = q = 1$ and $p = q = \infty$, where the wavelet-type systems are used to prove the upper bounds, were obtained by D. Bazarkhanov (see [23, 24]). D. Bazarkhanov also obtained right orders of the Fourier widths for the Triebel-Lizorkin classes.

Results of Subsection 4.4 on the orthowidths show that if we want to approximate classes \mathbf{W}_p^r or \mathbf{H}_p^r in L_q with (p, q) distinct from $(1, 1)$ and (∞, ∞) by operators of orthogonal projection of rank $m = |Q_n|$, then the best (in the sense of order) operator is S_{Q_n} . The operator S_{Q_n} is a very natural operator for approximating classes \mathbf{W}_p^r and \mathbf{H}_p^r . Therefore, we can ask the following question: How much can we weaken the assumption that the rank $m = |Q_n|$ linear operator is an orthogonal projection and still get that S_{Q_n} is the best (in the sense of order)? Here is a result in this direction. In [352] along with the quantities $\varphi_m(\mathbf{F}, L_p)$ we consider the quantities

$$\varphi_m^B(\mathbf{F}, L_q) = \inf_{G \in \mathcal{L}_m(B)_q} \sup_{f \in \mathbf{F} \cap D(G)} \|f - G(f)\|_q,$$

where $B \geq 1$ is a number and $\mathcal{L}_m(B)_q$ is the set of linear operators G with domains $D(G)$ containing all trigonometric polynomials, and with ranges contained in an m -dimensional subspace of L_q , such that $\|Ge^{i(\mathbf{k}, \mathbf{x})}\|_2 \leq B$ for all \mathbf{k} . It is clear that $\mathcal{L}_m(1)_2$ contains the operators of orthogonal projection onto m -dimensional subspaces, as well as operators given by multipliers $\{\lambda_l\}$ with $|\lambda_l| \leq 1$ for all l with respect to an orthonormal system of functions. It is known (see [352]) that in the case (p, q) distinct from $(1, 1)$ and (∞, ∞) S_{Q_n} gives the order of $\varphi_m^B(\mathbf{F}, L_q)$ for both $\mathbf{F} = \mathbf{W}_p^r$ and $\mathbf{F} = \mathbf{H}_p^r$. Another result of that same flavor is about nuclear widths mentioned above (see [139]).

4.5 The linear widths

As in the previous subsections, we begin with results on classes \mathbf{W}_p^r .

Theorem 4.45. *Let $r > (1/p - 1/q)_+$, $1 < p < \infty$ and $1 \leq q < \infty$. Then we have*

$$\lambda_m(\mathbf{W}_p^r, L_q) \asymp \begin{cases} \left(\frac{(\log m)^{d-1}}{m} \right)^{r-(1/p-1/q)_+}, & \text{for } q \leq 2, \text{ or } p \geq 2; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/2}, & \text{for } 1/p+1/q \geq 1, q > 2, r > 1/p; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/2+1/q}, & \text{for } 1/p+1/q \leq 1, p < 2, r > 1-1/q. \end{cases} \quad (4.5)$$

Theorem 4.45 for $1 < q < \infty$ was proved by Galeev [137, 140], and for $q = 1$ by Romanuk [277]. It is interesting to notice that by putting $M = \frac{m}{(\log m)^{d-1}}$, the relations (4.5) look like the asymptotic order of $\lambda_m(W_p^r, L_q)$ in the univariate case ($d = 1$), see e.g. [118].

We summarize the above results on the $\lambda_m(\mathbf{W}_p^r, L_q)$ in the following picture.

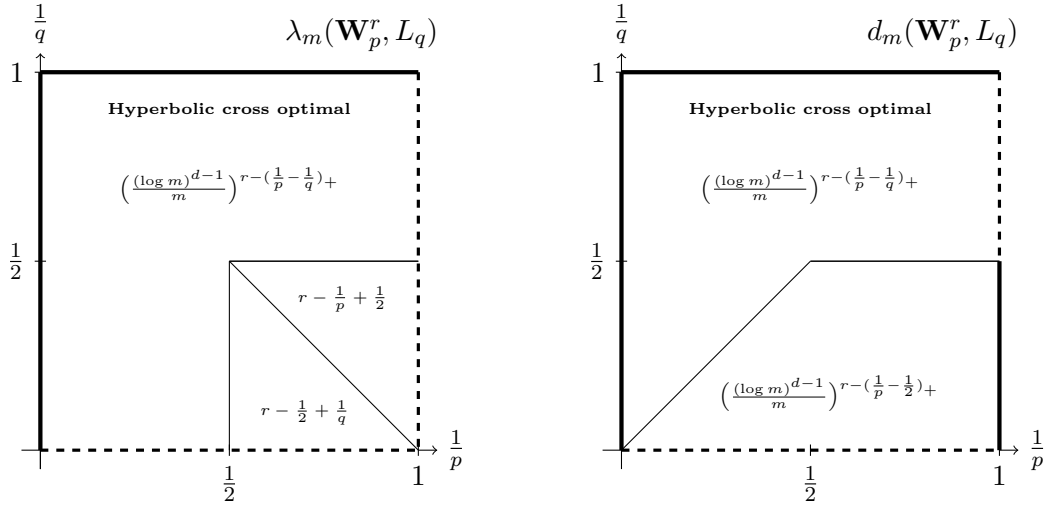


Figure 6: Comparison of $\lambda_m(\mathbf{W}_p^r, L_q)$ and $d_m(\mathbf{W}_p^r, L_q)$

The region

$$R_1 := \{(1/p, 1/q) : 0 < 1/p \leq 1/2, \quad \text{or} \quad 1/2 \leq 1/q \leq 1\}$$

is covered by Theorem 4.45. For this region the upper bounds follow from Theorem 4.4 and the fact that the operators S_{Q_n} are uniformly bounded from L_q to L_q , $1 < q < \infty$. It means that in this case the subspaces $\mathcal{T}(Q_n)$ of the hyperbolic cross polynomials are optimal in the sense of order. Theorem 4.45 shows that for the small square $(1/2, 1) \times (0, 1/2)$ the subspaces $\mathcal{T}(Q_n)$ are not optimal in the sense of order. Theorem 4.45 gives the orders of the $\lambda_m(\mathbf{W}_p^r, L_q)$ for all $(1/p, 1/q)$ from the square $(0, 1) \times (0, 1]$ under some restrictions on r . In all other cases the right orders of $\lambda_m(\mathbf{W}_p^r, L_q)$ are not known (see Open problem 4.8).

Theorem 4.46. *Let $1 \leq p, q \leq \infty$ and $r > (1/p - 1/q)_+$. Then we have*

$$\lambda_m(\mathbf{H}_p^r, L_q) \asymp \begin{cases} \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(d-1)/2}, & \text{for } 1 \leq q \leq p \leq \infty, \\ & p \geq 2, q < \infty; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/q} (\log m)^{(d-1)/q}, & \text{for } 1 < p \leq q \leq 2; \\ & 2 \leq p < q < \infty; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/2} (\log m)^{(d-1)/2}, & \text{for } 1/p + 1/q \geq 1, \\ & 2 < q < \infty, r > 1/p; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/2+1/q} (\log m)^{(d-1)/q}, & \text{for } 1/p + 1/q < 1, q < \infty, \\ & 1 < p \leq 2, r > 1 - 1/q. \end{cases}$$

The first line is due to Temlyakov [344]. We make a remark on this result, which is Theorem 3.2 from [344]. The most important part of this result is the lower bound for $\lambda_m(\mathbf{H}_\infty^r, L_1)$, which is proved in [344]. For the upper bounds in Theorem 3.2 from [344] the reference is given.

$$\lambda_m(\mathbf{H}_\infty^r, L_\infty) \asymp m^{-r}(\log m)^{r+1}.$$

Comments on the lower bound are given after Theorem 4.26 above. Approximation by V_{Q_n} with an appropriate n implies the upper bound. The right order of $\lambda_m(\mathbf{H}_\infty^r, L_\infty)$ in case $d \geq 3$ is not known. The cases $1 < p \leq q \leq 2$ as well as $1/p + 1/q \geq 1$, $2 < q < \infty$, $r > 1/p$ in Theorem 4.46 are due to Galeev [140]. The remaining cases have been settled very recently by Malykhin and Ryutin [229]. These authors provided sharp lower bounds (based on Lemma 4.34 above) to well-known upper bounds given by Galeev [140]. Note, that the case $1 \leq q \leq p < 2$ is still open. For an upper bound we refer the reader to the third case in Theorem 4.6 above.

Figure 1 consists of two plots showing the optimal solution for the HC problem. The left plot is for the $\lambda_m(\mathbf{H}_p^r, L_q)$ problem, and the right plot is for the $d_m(\mathbf{H}_p^r, L_q)$ problem. Both plots show the feasible region in the (α, β) plane, with the optimal solution marked by a question mark and labeled "HC opt."

Left Plot ($\lambda_m(\mathbf{H}_p^r, L_q)$): The feasible region is a square with vertices at $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$. The axes are labeled α and β , with tick marks at $0, \frac{1}{2}, 1$. The optimal solution is marked by a question mark and labeled "HC opt." in the region where $\alpha = r - \frac{1}{p} + \frac{1}{q}$ and $\beta = \frac{1}{q}$. Other regions are labeled "HC opt." with various α and β values.

Right Plot ($d_m(\mathbf{H}_p^r, L_q)$): The feasible region is a square with vertices at $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$. The axes are labeled α and β , with tick marks at $0, \frac{1}{2}, 1$. The optimal solution is marked by a question mark and labeled "HC opt." in the region where $\alpha = r - \left(\frac{1}{p} - \max\left\{\frac{1}{2}, \frac{1}{q}\right\}\right)_+$ and $\beta = \max\left\{\frac{1}{2}, \frac{1}{q}\right\}$. Other regions are labeled "HC opt." with various α and β values.

where α and β refers to the asymptotic order

The region

is covered by Theorem 4.46. For this region the upper bounds follow from Theorem 4.4, which means that in this case the subspaces $\mathcal{T}(Q_n)$ of the hyperbolic cross polynomials are optimal in the sense of order. The region

is covered by Theorem 4.46. Theorem 4.46 shows that for this region the subspaces $\mathcal{T}(Q_n)$ are not optimal in the sense of order. Theorem 4.46 gives the orders of the $\lambda_m(\mathbf{H}_p^r, L_q)$ for all $(1/p, 1/q)$ from the regions R_1 and R_2 under some restrictions on r . In all other cases the right orders of $\lambda_m(\mathbf{H}_p^r, L_q)$ are not known (see Open problem 4.8).

Theorem 4.47. *Let $r > (1/p - 1/q)_+$, $1 < p, q < \infty$ and $1 \leq \theta < \infty$. Then we have*

$$\lambda_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp \begin{cases} \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(1/2-1/\theta)_+(d-1)}, & \text{for } 2 \leq q < p; \\ & q \leq 2 \leq p, \theta \geq 2; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/q}, & \text{for } 2 \leq p < q, \\ & 2 \leq \theta \leq q, r > 1 - 1/q; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/q} (\log m)^{(1/q-1/\theta)_+(d-1)}, & \text{for } 1 < p < q \leq 2; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/2} (\log m)^{(1/2-1/\theta)_+(d-1)}, & \text{for } 1/p + 1/q \geq 1, q \geq 2, \\ & r > 1/p; \\ \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/2+1/q}, & \text{for } 1/p + 1/q \leq 1, 1 < p \leq 2, \\ & 2 \leq \theta \leq q, r > 1 - 1/q. \end{cases}$$

It is too complicated to summarize the above results on the $\lambda_m(\mathbf{B}_{p,\theta}^r, L_q)$ in a picture, since there are three parameters p, θ, q (with fixed r) which require a three-dimensional picture.

Theorem 4.48. *Let $r > 0$, $2 \leq p \leq \infty$ and $2 \leq \theta < \infty$. Then we have*

$$\lambda_m(\mathbf{B}_{p,\theta}^r, L_1) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r (\log m)^{(1/2-1/\theta)(d-1)}.$$

Theorem 4.49. *Let $r > 1 - 1/q$, $1 < q \leq 2$ and $1 \leq \theta \leq q$. Then we have*

$$\lambda_m(\mathbf{B}_{1,\theta}^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1+1/q} (\log m)^{(1/q-1/\theta)(d-1)}.$$

Theorem 4.50. *Let $r > 1/2$. Then we have*

$$\lambda_m(\mathbf{B}_{\infty,1}^r, L_\infty) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^r.$$

In Theorems 4.47 – 4.50, the case $1/p + 1/q \geq 1$, $2 \leq q < \infty$ was proved in [270], the case $q = 1$, $p \geq 2$ in [277], the other cases were proved in [272].

Let us give a brief comment on the proofs of Theorems 4.45 – 4.50. For details, the reader can see [140], [270], [272], [277]. For the cases $p \geq 2$ or $q \leq 2$, the upper bounds are derived from the linear approximation by the operators S_{Q_n} , and the lower bounds from the inequality (4.1) and corresponding lower bounds for Kolmogorov n -widths in Theorems 4.13 – 4.36. For the case $p \leq 2 < q$ the upper bounds can be reduced to the upper bounds of linear n -widths of finite-dimensional sets which can be estimated by the following Gluskin's lemma [144].

Lemma 4.51. *Let $1 \leq p < 2 \leq q < \infty$ and $1/p + 1/q \geq 1$. Then we have for $n > m$,*

$$\lambda_m(B_p^n, \ell_q^n) \asymp \max \{ m^{1/q-1/p}, (1 - n/m)^{1/2} \min \{ 1, m^{1/q} n^{-1/2} \} \}.$$

The lower bounds for the case $1/p + 1/q \geq 1$, $q \geq 2$ can be reduced by the norm inequality $\| \cdot \|_q \geq \| \cdot \|_2$ to the case $p \leq 2$, $q = 2$, for the case $1/p + 1/q \leq 1$, $p \leq 2$ to the upper bounds of linear n -widths of finite-dimensional sets which can be estimated by Lemma 4.51.

There are cases in which the right order of $\lambda_m(\mathbf{B}_{p,\theta}^r, L_q)$ are not known.

4.6 Open problems

We presented historical comments and a discussion of results, including open problems, in the above text of Section 4. We summarize here the most important comments on open problems. A number of asymptotic characteristics is discussed in this section: the Kolmogorov widths, the linear widths, and the orthonormal widths (the Fourier widths). It seems like the most complete results are obtained for the orthonormal widths (see Subsection 4.4). However, even in the case of orthonormal widths there are still unresolved problems. We mention the one for the \mathbf{W} classes.

Open problem 4.1. Find the order of $\varphi_m(\mathbf{W}_1^r, L_\infty)$ for all $r > 1$ and d .

Results of Subsection 4.3 show that the right order of the Kolmogorov widths $d_m(\mathbf{W}_p^r, L_q)$ are known for all $1 < p, q < \infty$ and $r > r(p, q)$. However, in the case of extreme values of p or q (p or q takes a value 1 or ∞) not much is known, see also Figure 4. Here are some open problems.

Open problem 4.2. Find the order of $d_m(\mathbf{W}_p^r, L_\infty)$ and $d_m(\mathbf{H}_p^r, L_\infty)$ for $2 \leq p \leq \infty$ and $r > 1/p$ in dimension $d \geq 3$.

Open problem 4.3. Find the order of $d_m(\mathbf{W}_p^r, L_\infty)$ for $1 \leq p < 2$.

Open problem 4.4 Find the order of $d_m(\mathbf{W}_1^r, L_q)$ for $1 \leq q < 2$.

It turns out that the problem of the right orders of the Kolmogorov widths for the \mathbf{H} classes is more difficult than this problem for the \mathbf{W} classes. In addition to some open problems in the case of extreme values of p and q the following case is not settled.

Open problem 4.5. Find the order of $d_m(\mathbf{H}_p^r, L_q)$ for $1 \leq q < p < 2$, see also Figure 5.

For the linear widths we have a picture similar to that for the Kolmogorov widths. We point out that the study of approximation of functions with mixed smoothness in the uniform norm (L_∞ norm) is very hard. Even the right orders of approximation by polynomials of special form – the hyperbolic cross polynomials – are not known in this case.

Open problem 4.6. Find the order of $E_{Q_n}(\mathbf{W}_q^r)_\infty$ for $2 < q \leq \infty$.

Open problem 4.7. Find the order of $E_{Q_n}(\mathbf{H}_\infty^r)_\infty$ for $d \geq 3$.

Open Problem 4.8. Find the right order of $\lambda_m(\mathbf{H}_p^r, L_q)$ in the missing cases in Figure 7.

5 Sampling recovery

In Section 4 we discussed approximation of functions with mixed smoothness by elements of finite dimensional subspaces. On the base of three asymptotic characteristics we concluded that subspaces $\mathcal{T}(Q_n)$ of the hyperbolic cross polynomials are optimal in the sense of order in many situations. The Kolmogorov width $d_m(\mathbf{F}, Y)$ gives the lower bound for approximation from any m -dimensional linear subspace. The linear width $\lambda_m(\mathbf{F}, Y)$ gives the lower bound for approximation from any m -dimensional linear subspace by linear operators. Finally, the orthonormal width $\varphi_m(\mathbf{F}, Y)$ gives the lower bound for approximation from any m -dimensional linear subspace by operators of orthogonal projections (assuming that the setting makes sense). In addition, from an applied point of view, restrictions on approximation methods in a form of linear operator and orthogonal projection, there is one more natural setting. In this setting we approximate (as above) by elements from finite dimensional subspaces but our methods of approximation are restricted to linear methods, which may only use the function values at a certain set of points. We discuss this setting in detail in this section. We begin with precise definitions. The general goal of this section is to recover a multivariate continuous periodic function $f : \mathbb{T}^d \rightarrow \mathbb{C}$ belonging to a function class \mathbf{F} from a finite set of m function values. To be more precise, we consider linear reconstruction algorithms of type

$$\Psi_m(f, X_m) := \sum_{i=1}^m f(\mathbf{x}^i) \psi_i(\cdot)$$

for multivariate functions. The set of sampling nodes $X_m := \{\mathbf{x}^i\}_{i=1}^m \subset \mathbb{T}^d$ and associated (continuous) functions $\Psi_m := \{\psi_i\}_{i=1}^m$ is fixed in advance. To guarantee reasonable access to function values we need that the function class \mathbf{F} consists of functions or equivalence classes which have a continuous representative. When considering classes \mathbf{W}_p^r or $\mathbf{B}_{p,\theta}^r$ the embedding into $C(\mathbb{T}^d)$ holds whenever $r > 1/p$, see Lemma 3.12 above.

As usual we are interested in the minimal error (sampling numbers/widths) in a Banach space Y

$$\varrho_m(\mathbf{F}, Y) := \inf_{X_m} \inf_{\Psi_m} \Psi_m(\mathbf{F}, X_m)_Y, \quad (5.1)$$

where

$$\Psi_m(\mathbf{F}, X_m)_Y := \sup_{f \in \mathbf{F}} \|f - \Psi_m(f, X_m)\|_Y.$$

The quantities $\varrho_m(\mathbf{F}, Y)$ are called *sampling widths* or *sampling numbers*. It is clear that the following inequalities hold

$$d_m(\mathbf{F}, Y) \leq \lambda_m(\mathbf{F}, Y) \leq \varrho_m(\mathbf{F}, Y). \quad (5.2)$$

There are no general inequalities relating the characteristics $\varphi_m(\mathbf{F}, Y)$ and $\varrho_m(\mathbf{F}, Y)$. In both cases in addition to the linearity assumption on the approximation operator we impose additional restrictions but those restrictions are of a very different nature – orthogonal projections and sampling operators. However, it turns out that similarly to the orthonormal setting, where $\mathcal{T}(Q_n)$ are optimal (in the sense of order) in all cases with a few exceptions, the subspaces $\mathcal{T}(Q_n)$ are optimal (in the sense of order) from the point of view of $\varrho_m(\mathbf{F}, Y)$ in all cases where we know their right orders.

We make a more detailed comment on this issue on the example of the univariate problem. Classically, the interpolation problem by polynomials (algebraic and trigonometric) was only considered in the space of continuous functions and the error of approximation was measured

in the uniform norm (here, for notational convenience, we denote it L_∞ norm). Restriction to continuous functions is very natural because we need point evaluations of the function in the definition of the recovery operator Ψ_m . We are interested in the recovery error estimates not only in the uniform norm L_∞ , but in the whole range of L_q , $1 \leq q \leq \infty$. It was understood in the first papers on this topic in the mid of 1980s (see, for instance, [335]) that recovery in L_q , $q < \infty$, instead of L_∞ brings difficulties and new phenomena. However, the problem of optimal recovery on classes W_p^r and H_p^r in L_q was solved for all $1 \leq p, q \leq \infty$, $r > 1/p$ (see [352]):

$$\varrho_m(W_p^r, L_q) \asymp \varrho_m(H_p^r, L_q) \asymp m^{-r+(1/p-1/q)_+}.$$

Thus, in the univariate case the asymptotic characteristics ϱ_m and φ_m behave similarly and the optimal subspaces for recovery ($1 \leq p, q \leq \infty$) and orthowidth ($1 \leq p, q \leq \infty$, $(p, q) \neq (1, 1)$, $(p, q) \neq (\infty, \infty)$) are the trigonometric polynomials.

It was established in Section 4 that in some cases the right order of the linear width can be realized by an orthogonal projection operator and in other cases it cannot be realized that way. It means that in the first case $\lambda_m(\mathbf{F}, Y) \asymp \varphi_m(\mathbf{F}, Y)$ and in the second case $\lambda_m(\mathbf{F}, Y) = o(\varphi_m(\mathbf{F}, Y))$. It is an interesting problem: When the linear width can be realized (in the sense of order) by a sampling operator? In other words: When $\lambda_m(\mathbf{F}, Y) \asymp \varrho_m(\mathbf{F}, Y)$? Note, that in the univariate setting an optimal algorithm in the sense of order of λ_m for the classes W_p^r in L_p , $1 < p < \infty$, $r > 1/p$, consists in the standard equidistant interpolation method. In that sense, λ_m and ϱ_m are equal in order. However, when considering multivariate classes the notion of “equidistant” is not clear anymore. In other words, what are optimal point sets X_m in the d -dimensional cube to sample the function and build optimal sampling algorithms in the sense of order of ϱ_m ? Is such an operator then also optimal in the sense of order of λ_m ? For the classes \mathbf{W}_p^r this represents a well-known open problem.

Let us begin our discussion of known results with Hölder-Nikol’skii classes \mathbf{H}_p^r . Temlyakov [326] proved for $1 \leq p \leq \infty$ and $r > 1/p$ the relation

$$\varrho_m(\mathbf{H}_p^r, L_p) \lesssim m^{-r}(\log m)^{(r+1)(d-1)}. \quad (5.3)$$

The correct order is not known here, except in the situation $d = 2$, $p = \infty$, and $r > 1/2$, see Remark 5.14 below. Let the Smolyak-type sampling operators T_n (see Subsection 4.2) be induced by the univariate linear operators

$$Y_s(f) := R_{2^s}(f) := 2^{-s-2} \sum_{l=1}^{2^{s+2}} f(x(l)) \mathcal{V}_{2^s}(x - x(l)), \quad x(l) := \pi l 2^{-s-1}.$$

Then (5.3) follows from the error bound [326]

$$\sup_{f \in \mathbf{H}_p^r} \|f - T_n(f)\|_p \lesssim 2^{-rn} n^{d-1}. \quad (5.4)$$

Note that with the above specification of Y_s we have $\Delta_s(f) \in \mathcal{T}(2^{s+1})$ and $T_n(f) \in \mathcal{T}(Q_{n+d})$. Comparing (5.4) with Theorem 4.6 we see that the above sampling operator T_n with $m \asymp |Q_n|$ does not provide the order of best approximation $E_{Q_n}(\mathbf{H}_p^r)_p$ for $1 < p < \infty$. We note that the first result on recovering by sampling operators of the type of T_n (Smolyak-type operators) was obtained by Smolyak [305]. He proved the error bound

$$\sup_{f \in \mathbf{W}_\infty^r} \|f - T_n(f)\|_\infty \lesssim 2^{-rn} n^{d-1}. \quad (5.5)$$

The error bound (5.5) and similar bound for $1 \leq p < \infty$ follow from (5.4) by embedding, see (3.13).

The first correct order

$$\varrho_m(\mathbf{H}_p^r, L_q) \asymp m^{-(r-1/p+1/q)} (\log m)^{(r-1/p+2/q)(d-1)}, \quad m \in \mathbb{N}, \quad (5.6)$$

for $1 < p < q \leq 2$, $r > 1/p$, was proven by Dinh Dũng [91]. The upper bound is given by the above sampling operator T_n with $m \asymp |Q_n|$: for $1 \leq p < q < \infty$,

$$\sup_{f \in \mathbf{H}_p^r} \|f - T_n(f)\|_q \lesssim 2^{-(r-1/p+1/q)n} n^{(d-1)/q}. \quad (5.7)$$

The proof is based on the case $p = q$, discussed above, and Remark 2.12 to Theorem 2.11 (see inequality (2.22)). Indeed, $f \in \mathbf{H}_p^r$ implies $\|v_j(f)\|_p \lesssim 2^{-r|j|_1}$ (see (5.16) below). Application of inequality (2.22) completes the proof. The lower bound for $1 < p < q \leq 2$ follows from the inequality $\varrho_m \geq d_m$ and the lower bound of $d_m(\mathbf{H}_p^r, L_q)$ for $1 < p < q \leq 2$ obtained by Galeev [138]. Comparing (5.7) with Theorem 4.6 we see that the sampling operator T_n provides the best order of approximation $E_{Q_n}(\mathbf{H}_p^r)_q$ in the case $1 \leq p < q < \infty$.

Surprisingly, even in the Hilbert space setting, i.e., for Sobolev classes \mathbf{W}_2^r there are only partial results for $\varrho_m(\mathbf{W}_2^r, L_2)$. Let us mention the following result due to Temlyakov [351]. The following situation deals with the error norm L_∞ and provides a sharp result. For $r > 1/2$ we have

$$\varrho_m(\mathbf{W}_2^r, L_\infty) \asymp m^{-(r-1/2)} (\log m)^{r(d-1)}. \quad (5.8)$$

Interestingly, we have here $\varrho_m \asymp \lambda_m$. The difficult part of (5.8) is the upper bound. Its proof uses Theorem 2.11. The lower bounds for λ_m were reduced in [351] to known lower bounds for $d_m(\mathbf{W}_1^r, L_2)$ using the Ismagilov duality result for the λ_m [181]. Notice that using properties of 2-summing operators the corresponding lower bound can be derived from a result proven recently by Cobos, Kühn, Sickel [59]. They showed the beautiful identity

$$\lambda_m(\mathbf{W}_2^r, L_\infty) = \left(\sum_{j=m}^{\infty} \lambda_j(\mathbf{W}_2^r, L_2)^2 \right)^{1/2}, \quad (5.9)$$

which immediately gives the lower bound in (5.8).

The known results on the hyperbolic cross approximation (see Theorem 3.7 from [352], Chapter 3) imply: for $1 < p < \infty$, $r > 1/p$

$$E_{Q_n}(\mathbf{W}_p^r)_\infty = o \left(\sup_{f \in \mathbf{W}_p^r} \|f - T_n(f)\|_\infty \right).$$

Similarly to the class \mathbf{H}_p^r , there holds the correct order relation

$$\varrho_m(\mathbf{W}_p^r, L_q) \asymp m^{-(r-1/p+1/q)} (\log m)^{(r-1/p+1/q)(d-1)} \quad (5.10)$$

if either $1 < p < q \leq 2$ or $2 \leq p < q < \infty$ and $r > 1/p$, see Theorem 5.12(i)-(ii) below which has been proved recently in [48] for $r > 1/p$, and in [103] for $r > \max\{1/p, 1/2\}$. The special case $p = 2 < q$ is proved in [47, Thm. 6.10]. If we replace the uniform error norm by the L_p -error in (5.8) we can only say the following for $r > \max\{1/p, 1/2\}$ (including $p = 2$), namely

$$\varrho_m(\mathbf{W}_p^r, L_p) \lesssim m^{-r} (\log m)^{(r+1/2)(d-1)}. \quad (5.11)$$

This result has been first observed by Sickel [296] for the case $d = 2$. It was extended later to arbitrary d by Sickel, Ullrich [298], [391]. A corresponding lower bound is not known. Note, that the linear widths in this situation are smaller than the right-hand side of (5.11). In fact, there we have $(\log m)^{(d-1)r}$ instead of $(\log m)^{(r+1/2)(d-1)}$. Let us mention once more that even in the case $p = 2$ it is not known

- whether the upper bound in (5.11) is sharp and
- whether the sampling widths ϱ_m coincide with the linear widths λ_m in the sense of order.

It is remarkable to notice that we so far only have sharp bounds for the order of $\varrho_m(\mathbf{W}_p^r, L_q)$ and $\varrho_m(\mathbf{H}_p^r, L_q)$ in case $p < q$ and for these cases the Smolyak algorithm is optimal in the sense of order.

In general we do not assume that the approximant $\Psi_m(f, X_m)$ on f satisfies the interpolation property

$$\Psi_m(f, X_m)(\mathbf{x}^i) = f(\mathbf{x}^i) \quad , \quad i = 1, \dots, m. \quad (5.12)$$

The already described upper bounds for the sampling widths in the various situations are based on sampling algorithms on sparse grids obtained by applying the Smolyak algorithm to univariate sampling operators, see Subsection 4.2 above. In the sequel we will describe how to create and analyze such operators. As an ingredient we may use the univariate classical de la Vallée Poussin interpolation which is described in the next subsection. The second step is a tensorization procedure. Afterwards we are going to prove a characterization of the spaces of interest in terms of those sampling operators which finally yield the stated error bounds.

5.1 The univariate de la Vallée Poussin interpolation

For $m \in \mathbb{N}$ let $\mathcal{V}_m := \mathcal{V}_{m,2m}$ be the univariate de la Vallée Poussin kernel as introduced in Paragraph 2.1.3. An elementary calculation, see Paragraph 2.1.3, shows that

$$\mathcal{V}_m(t) := \frac{1}{m} \sum_{k=m}^{2m-1} \mathcal{D}_k(t) = \frac{\sin(mt/2) \sin(3mt/2)}{m \sin^2(t/2)} \quad , \quad m \in \mathbb{N}.$$

In the manner of (2.10) we define for $f \in C(\mathbb{T})$ the interpolation operator

$$V_m(f, J_{3m}) := \frac{1}{3m} \sum_{\ell=0}^{3m-1} f\left(\frac{2\pi\ell}{3m}\right) \mathcal{V}_m\left(\cdot - \frac{2\pi\ell}{3m}\right) \quad , \quad m \in \mathbb{N}, \quad (5.13)$$

with respect to the equidistant grid

$$J_N := \left\{ \frac{2\pi\ell}{N} : \ell = 0, \dots, N-1 \right\} \subset \mathbb{T} \quad , \quad N = 3m.$$

Instead of J_{3m} one can also use more redundant grids in the definition of (5.13) like, e.g., J_{4m} in [352] or J_{8m} in [326]. Note, that the particular choice J_{3m} in (5.13) leads to $\mathcal{V}_m(0) = 3m$ and $\mathcal{V}_m(2\pi\ell/3m) = 0$ if $\ell \neq 0$. Using that, (5.13) implies the interpolation property (5.12). It is straight-forward to compute the Fourier coefficients of the approximant $V_m(f, J_{3m})$. This directly leads to the following crucial reproduction result similar to Theorem 2.2.

Lemma 5.1. *Let $m \in \mathbb{N}$. The operator $V_m(\cdot, J_{3m})$ does not change trigonometric polynomials of degree at most m , i.e., if $f \in \mathcal{T}(m)$ is a trigonometric polynomial then*

$$V_m(f, J_{3m}) = f.$$

Let us also recall the classical trigonometric interpolation from (2.4) by

$$D_m(f, J_{2m+1}) := \frac{1}{2m+1} \sum_{\ell=0}^{2m} f\left(\frac{2\pi\ell}{2m+1}\right) \mathcal{D}_m\left(\cdot - \frac{2\pi\ell}{2m+1}\right) \quad , \quad m \in \mathbb{N}.$$

This approximant is also interpolating its argument in the nodes J_{2m+1} and Lemma 5.1 keeps valid also for this interpolation operator. However, it has two disadvantages. The first one is the fact that subsequent dyadic grids are not nested, i.e., when comparing $D_{2^{j+1}}$ and D_{2^j} . However, this can be fixed by using the modified Dirichlet kernel \mathcal{D}_n^1 , see (2.6). The second issue is related to the effects described in Theorems 2.1 and 2.2. In the case of extreme value of p or q it might be a problem to work with the Dirichlet kernels. In fact, it is known that the classical trigonometric interpolation does not provide the optimal rate of recovery in case $p = 1$ and $p = \infty$. That might be no problem when considering Sobolev classes \mathbf{W}_p^r for $1 < p < \infty$. However, the cases $p = 1$ and $p = \infty$ represent important special cases for Besov classes $\mathbf{B}_{p,\theta}^r$. All the mentioned de la Vallée Poussin sampling operators show the following behavior in the univariate setting, see for instance [352], [297] and [389] for further details.

Theorem 5.2. (i) *Let $1 \leq p, \theta \leq \infty$ and $r > 1/p$ then*

$$\|f - V_m(f, J_{3m})\|_p \lesssim m^{-r} \|f\|_{\mathbf{B}_{p,\theta}^r}.$$

(ii) *Let $1 < p < \infty$ and $r > 1/p$ then*

$$\|f - V_m(f, J_{3m})\|_p \lesssim m^{-r} \|f\|_{\mathbf{W}_p^r}.$$

Let us emphasize that the use of a sampling operator in L_p , $p < \infty$, immediately brings problems. For instance, it is easy to prove that

$$\|V_m(f, J_{3m})\|_\infty \leq C \|f\|_\infty.$$

It is also easy to understand that we do not have an analog of the above inequality in the L_p spaces with $p < \infty$. It was understood in early papers on this topic (see [326]) that the technically convenient way out of the above problem is to consider a superposition of a sampling operator and the de la Vallée Poussin operator. In particular, the following inequality was established (see Lemma 6.2 from [352], Chapter 1, and Corollary 3 from [91])

$$\|V_m(V_n(f), J_{3m})\|_p \leq C \|f\|_p (n/m)^{1/p}, \quad n \geq m. \quad (5.14)$$

5.2 Frequency-limited sampling representations - discrete Littlewood-Paley

The definition of the \mathbf{H} , \mathbf{B} and \mathbf{W} classes in terms of $\{\delta_{\mathbf{s}}(f)\}_{\mathbf{s}}$ is convenient for analyzing performance of the operators S_{Q_n} and not very convenient for analyzing operators T_n . The idea is to replace the convolutions $\delta_{\mathbf{s}}(f)$ in (3.9), (3.5) in the definition of the Besov and Sobolev space by discrete convolutions of type (5.13). Having such a sampling representation at hand one can easily obtain bounds for the sampling approximation error as we will show below.

In order to obtain proper characterizations of $\mathbf{B}_{p,\theta}^r$ we need the following dyadic differences of (5.13) given by

$$v_0(f) := V_1(f, J_3) \quad , \quad v_j(f) := V_{2^j}(f, J_{3 \cdot 2^j}) - V_{2^{j-1}}(f, J_{3 \cdot 2^{j-1}}) \quad , \quad j \geq 1.$$

In other words we set $Y_j := V_{2^j}(\cdot, J_{3 \cdot 2^j})$ from the Smolyak-type algorithm construction in Subsection 4.2. For the d -variate situation we need their tensor product $\Delta_{\mathbf{j}} := v_{\mathbf{j}} := \bigotimes_{n=1}^d v_{j_n}$.

This operator has to be understood componentwise. For $f \in C(\mathbb{T}^d)$ the function $(\bigotimes_{i=1}^d v_{j_n})f$ is the trigonometric polynomial from $\mathcal{T}(2^{\mathbf{j}+1}, d)$ which we obtain by applying each v_{j_n} to the respective component of f . This gives the representation

$$f = \sum_{\mathbf{j} \in \mathbb{N}_0^d} v_{\mathbf{j}}(f). \quad (5.15)$$

In case of \mathbf{H}_p^r the inequality

$$\sup_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1} \|v_{\mathbf{j}}(f)\|_p \lesssim \|f\|_{\mathbf{H}_p^r} \quad (5.16)$$

is due to Temlyakov [326]. The inverse inequality (see [91])

$$\|f\|_{\mathbf{H}_p^r} \lesssim \sup_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1} \|v_{\mathbf{j}}(f)\|_p \quad (5.17)$$

follows from a simple observation: suppose that

$$f = \sum_{\mathbf{j}} t_{\mathbf{j}}, \quad \|t_{\mathbf{j}}\|_p \leq 2^{-r|\mathbf{j}|_1}, \quad t_{\mathbf{j}} \in \mathcal{T}(2^{\mathbf{j}}, d),$$

then $\|f\|_{\mathbf{H}_p^r} \leq C(d)$.

The following proposition is a nontrivial generalization of inequalities (5.16) and (5.17) to the case of \mathbf{B} and \mathbf{W} classes.

Proposition 5.3. (i) Let $1 \leq p, \theta \leq \infty$ and $r > 1/p$. Then we have for any $f \in \mathbf{B}_{p,\theta}^r$

$$\left(\sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1 \theta} \|v_{\mathbf{j}}(f)\|_p^\theta \right)^{1/\theta} \asymp \|f\|_{\mathbf{B}_{p,\theta}^r} \quad (5.18)$$

with the sum being replaced by a supremum for $\theta = \infty$.

(ii) Let $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$ then we have for any $f \in C(\mathbb{T}^d)$

$$\left\| \left(\sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1^2} |v_{\mathbf{j}}(f)|^2 \right)^{1/2} \right\|_p \asymp \|f\|_{\mathbf{W}_p^r}. \quad (5.19)$$

The idea to replace the Littlewood-Paley convolutions $A_{\mathbf{j}}(f)$, see (2.14), in the description of Besov classes by dyadic blocks $v_{\mathbf{j}}(f)$ coming out of a discrete convolution goes back to Dinh Dũng [91, 96]. The relation (5.18) in Proposition 5.3 was proven in [96], and the relation (5.19) has been proven in the recent Preprint [48], see also [47] for the case $p = 2$.

Proof. For the convenience of the reader let us give a short proof of the upper bound in (5.18) to show the main techniques. This proof alike the proof of (5.3) is based on relation (5.14). We start with the decomposition $f = \sum_{\mathbf{m} \in \mathbb{N}_0^d} f_{\mathbf{m}}$, where $f_{\mathbf{m}}$ is given by (3.10). Due to the reproduction property in Lemma 5.1 we obtain

$$\|v_{\mathbf{j}}(f)\|_p \leq \sum_{\substack{m_n \geq j_n - 3 \\ n=1, \dots, d}} \|v_{\mathbf{j}}(f_{\mathbf{m}})\|_p. \quad (5.20)$$

Choose now $1/p < a < r$. Relation (5.14) yields

$$\|v_{\mathbf{j}}(f_{\mathbf{m}})\|_p \lesssim 2^{(|\mathbf{m}|_1 - |\mathbf{j}|_1)/p} \|f_{\mathbf{m}}\|_p \quad (5.21)$$

which gives

$$2^{|\mathbf{j}|_1/p} \|v_{\mathbf{j}}(f)\|_p \lesssim \sum_{\substack{m_n \geq j_n - 3 \\ n=1, \dots, d}} 2^{-|\mathbf{m}|_1(a-1/p)} 2^{|\mathbf{m}|_1 a} \|f_{\mathbf{m}}\|_p.$$

What remains is a standard argument based on a discrete Hardy type inequality, see [73, Lem. 2.3.4]. Hölder's inequality with $1/\theta + 1/\theta' = 1$ yields

$$2^{|\mathbf{j}|_1/p} \|v_{\mathbf{j}}(f)\|_p \lesssim 2^{-|\mathbf{j}|_1(a-1/p)} \left(\sum_{\substack{m_n \geq j_n - 3 \\ n=1, \dots, d}} 2^{|\mathbf{m}|_1 a \theta} \|f_{\mathbf{m}}\|_p^\theta \right)^{1/\theta}.$$

Taking the ℓ_θ -norm (with respect to \mathbf{j}) on both sides and interchanging the summation on the right-hand side yields

$$\begin{aligned} \sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{|\mathbf{j}|_1 r \theta} \|v_{\mathbf{j}}(f)\|_p^\theta &\lesssim \sum_{\mathbf{m}} 2^{|\mathbf{m}|_1 a \theta} \|f_{\mathbf{m}}\|_p^\theta \sum_{\substack{j_n \leq m_n + 3 \\ n=1, \dots, d}} 2^{|\mathbf{j}|_1(r-a)\theta} \\ &\lesssim \sum_{\mathbf{m}} 2^{|\mathbf{m}|_1 r \theta} \|f_{\mathbf{m}}\|_p^\theta. \end{aligned}$$

■

The relation in Proposition 5.3(ii) requires further tools from Fourier analysis, i.e., maximal functions of Peetre and Hardy-Littlewood type. The principle proof technique is the same. However, here we have to deal with pointwise estimates of $|v_{\mathbf{j}}(f_{\mathbf{m}})(\cdot)|$. This can be handled by using the pointwise decay property (2.8) above together with the Peetre and (later) Hardy-Littlewood maximal function. Applying the corresponding vector-valued maximal inequalities requires the condition $r > \max\{1/p, 1/2\}$. For details we refer to the recent Preprint [48] and [47].

The method in the proof of Proposition 5.3 can be adapted to more general Smolyak algorithms, see also Propositions 4.1, 4.2 above and [298].

In the following one-sided relation the condition on r can be relaxed to $r > 0$.

Proposition 5.4. *Let $\{t_{\mathbf{j}}\}_{\mathbf{j}}$ be a sequence of trigonometric polynomials with $t_{\mathbf{j}} \in \mathcal{T}(2^{\mathbf{j}}, d)$ such that the respective right-hand side (5.22) or (5.23) below is finite.*

(i) *Let $1 \leq p, \theta \leq \infty$ and $r > 0$. Then $f = \sum_{\mathbf{j} \in \mathbb{N}_0^d} t_{\mathbf{j}}$ belongs to $\mathbf{B}_{p,\theta}^r$ and*

$$\|f\|_{\mathbf{B}_{p,\theta}^r} \lesssim \left(\sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1 \theta} \|t_{\mathbf{j}}\|_p^\theta \right)^{1/\theta}. \quad (5.22)$$

(ii) *Let $1 < p < \infty$ and $r > 0$. Then $f = \sum_{\mathbf{j} \in \mathbb{N}_0^d} t_{\mathbf{j}}$ belongs to \mathbf{W}_p^r and*

$$\|f\|_{\mathbf{W}_p^r} \lesssim \left\| \left(\sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1 2} |t_{\mathbf{j}}(\cdot)|^2 \right)^{1/2} \right\|_p. \quad (5.23)$$

The proof is similar to the one of Proposition 5.3, for details see [47, 48].

5.3 Sampling on the Smolyak grids

It is now a standard way to obtain good errors of recovering functions with mixed smoothness by using the Smolyak-type algorithms described in Subsection 4.2. Different realizations differ by the family of operators $\{Y_s\}_{s=0}^\infty$ used in the construction. Smolyak [304] and other authors (see, for instance, [326, 91, 96]) used interpolation-type operators based on the Dirichlet or de la Vallée Poussin kernels.

For $n \in \mathbb{N}_0^d$ we define the sampling operator T_n by

$$T_n(f) := \sum_{\substack{\mathbf{j} \in \mathbb{N}_0^d \\ |\mathbf{j}|_1 \leq n}} v_{\mathbf{j}}(f) \quad , \quad f \in C(\mathbb{T}^d), \quad (5.24)$$

where $v_{\mathbf{j}}(f)$ are defined in Subsection 5.2.

From Lemma 5.1 we see that T_n does not change trigonometric polynomials from hyperbolic crosses Q_{n-3d} , see (2.15). In addition, T_n interpolates f at every grid point $\mathbf{y} \in \widetilde{SG}^d(n)$, i.e.,

$$T_n(f, \mathbf{y}) = f(\mathbf{y}), \quad \mathbf{y} \in \widetilde{SG}^d(n), \quad (5.25)$$

where the Smolyak Grid $\widetilde{SG}^d(n)$ of level n is given by

$$\widetilde{SG}^d(n) := \bigcup_{j_1 + \dots + j_d \leq n} J_{3 \cdot 2^{j_1}} \times \dots \times J_{3 \cdot 2^{j_d}}. \quad (5.26)$$

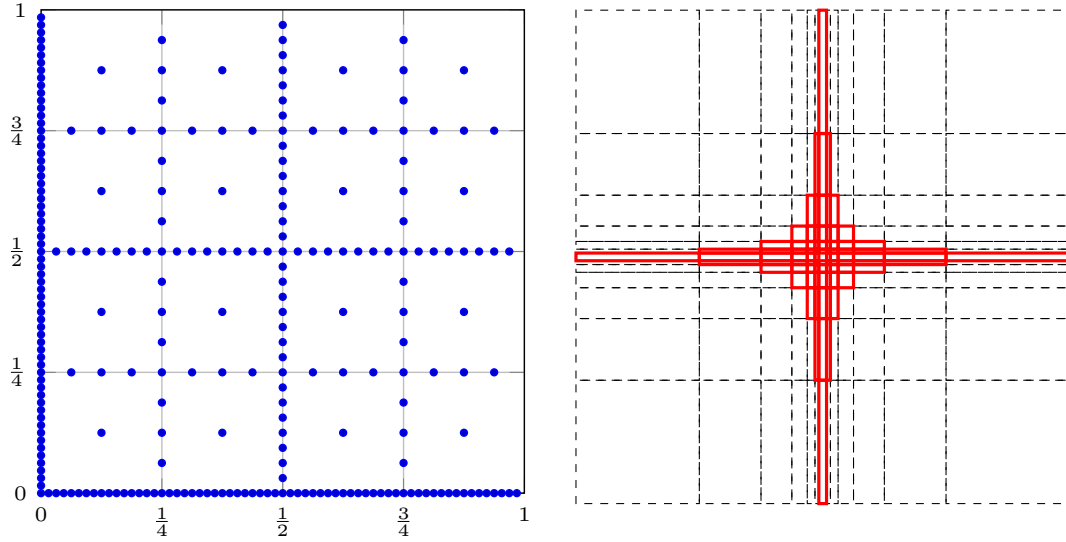


Figure 8: A sparse grid and associated hyperbolic cross in $d = 2$

The following estimate concerning the grid size is known, see for instance [45] or the recent paper [108, Lem. 3.10]. Clearly, $\text{card } \widetilde{SG}^d(n) = 3^d \text{card } SG^d(n)$, where $SG^d(n)$ is given in this section below (see also (8.4)). It holds

$$2^n \binom{n+d-1}{d-1} \leq \text{card } SG^d(n) \leq 2^{n+1} \binom{n+d-1}{d-1}. \quad (5.27)$$

This gives $\text{card } SG^d(n) \asymp 2^n n^{d-1}$. Here the nestedness of the univariate grids plays an important role. In fact, we can replace $j_1 + \dots + j_d \leq n$ by $j_1 + \dots + j_d = n$ in (5.26).

Let us begin with the recovery on the Smolyak grids of Sobolev classes \mathbf{W}_p^r in L_q .

Theorem 5.5. *Let $1 < p, q < \infty$.*

(i) *In case $p \geq q$ and $r > \max\{1/p, 1/2\}$ it holds*

$$\sup_{f \in \mathbf{W}_p^r} \|f - T_n(f)\|_q \asymp 2^{-nr} n^{(d-1)/2}.$$

(ii) *In case $1 < p < q < \infty$ and $r > 1/p$ we have*

$$\sup_{f \in \mathbf{W}_p^r} \|f - T_n(f)\|_q \asymp 2^{-n(r-1/p+1/q)}.$$

Proof. Let us demonstrate the general principle on the special case $1 < p = q < \infty$. Starting with (5.30) (see below) we estimate using Hölder's inequality

$$\begin{aligned} \|f - T_n(f)\|_p &\leq \left\| \sum_{|\mathbf{j}|_1 > n} |v_{\mathbf{j}}(f)(\cdot)| \right\|_p \\ &\lesssim \left\| \left(\sum_{|\mathbf{j}|_1 > n} 2^{-2r|\mathbf{j}|_1} \right)^{1/2} \cdot \left(\sum_{|\mathbf{j}|_1 > n} 2^{2r|\mathbf{j}|_1} |v_{\mathbf{j}}(f)(\cdot)|^2 \right)^{1/2} \right\|_p \end{aligned} \quad (5.28)$$

Proposition 5.3 together with (5.31) finishes the proof. ■

In case $q = \infty$ we observe an extra log-term. The result in Theorem 5.6 below has been obtained by Temlyakov, see [351]. For the convenience of the reader we give a short proof of this result based on the non-trivial embedding in Lemma 3.14, which in turn is a corollary of Theorem 2.11.

Theorem 5.6. *For $1 < p < \infty$ and $r > 1/p$ we have*

$$\sup_{f \in \mathbf{W}_p^r} \|f - T_n(f)\|_{\infty} \asymp 2^{-n(r-1/p)} n^{(d-1)(1-1/p)}.$$

Proof. The proof from [351] goes as follows. First, the upper bounds are proved. The key role in that proof is played by Theorem 2.11. Second, the lower bound in the case $p = 2$ is proved. It is derived from the lower bound for $\lambda_m(\mathbf{W}_2^r, L_{\infty})$. Finally, using the lower bound for $p = 2$ and the upper bounds for $1 < p < \infty$ we derive the lower bounds for $1 < p' < \infty$. We only illustrate how to prove the upper bounds. In the proof below instead of direct use of Theorem 2.11 we use its corollary in the form of Lemma 3.14.

We use the triangle inequality to obtain

$$\begin{aligned} \|f - T_n(f)\|_{\infty} &\leq \sum_{|\mathbf{j}|_1 > n} \|v_{\mathbf{j}}(f)\|_{\infty} \\ &= \sum_{|\mathbf{j}|_1 > n} 2^{-(r-1/p)|\mathbf{j}|_1} 2^{(r-1/p)|\mathbf{j}|_1} \|v_{\mathbf{j}}(f)\|_{\infty}. \end{aligned}$$

Using Hölder's inequality with respect to $1/p + 1/p' = 1$ we obtain

$$\|f - T_n(f)\|_{\infty} \lesssim \left(\sum_{|\mathbf{j}|_1 > n} 2^{-(r-1/p)|\mathbf{j}|_1 p'} \right)^{1/p'} \left(\sum_{|\mathbf{j}|_1 > n} 2^{p(r-1/p)|\mathbf{j}|_1} \|v_{\mathbf{j}}(f)\|_{\infty}^p \right)^{1/p}.$$

Applying (5.18) we have

$$\|f - T_n(f)\|_\infty \lesssim 2^{-n(r-1/p)} n^{(d-1)(1-1/p)} \|f\|_{\mathbf{B}_{\infty,p}^{r-1/p}}.$$

Finally, the embedding $\mathbf{W}_p^r \hookrightarrow \mathbf{B}_{\infty,p}^{r-1/p}$ (see Lemma 3.14) concludes the proof. \blacksquare

The results in [340] (see Theorems 5.4 and 3.1 of Chapter 2) imply the inequality

$$\sup_{f \in \mathbf{W}_p^r} \|f - U_n(f)\|_\infty \gtrsim 2^{-(r-1/p)n} n^{(d-1)(1-1/p)}, \quad 1 < p < \infty, \quad (5.29)$$

which is valid for any sequence of linear operators $U_n : \mathbf{W}_p^r \rightarrow \mathcal{T}(Q_n)$. Theorem 5.6 and relation (5.29) show that the sequence of operators T_n is optimal (in the sense of order) among all linear operators of approximating by means of polynomials in $\mathcal{T}(Q_{n+d})$.

Let us now proceed to the \mathbf{H} -classes.

Theorem 5.7. *Let $1 \leq p, q \leq \infty$ and $r > 1/p$. Then we have the following.*

(i) *For $p \geq q$,*

$$\sup_{f \in \mathbf{H}_p^r} \|f - T_n(f)\|_q \asymp 2^{-rn} n^{d-1}.$$

(ii) *For $p < q$,*

$$\sup_{f \in \mathbf{H}_p^r} \|f - T_n(f)\|_q \asymp \begin{cases} 2^{-(r-1/p+1/q)n} n^{(d-1)/q}, & q < \infty, \\ 2^{-(r-1/p)n} n^{d-1}, & q = \infty. \end{cases}$$

The upper bounds in Theorem 5.7 for the case $p \geq q$ are already in (5.4). The corresponding lower bounds follow from more general results on lower bounds for numerical integration (see, for instance, Theorem 8.6 below). In particular, Theorem 8.6 provides these lower bounds not only for the operator T_n but for any recovering operator, which uses the same nodes as T_n . In the case $p < q$ the lower bounds follow from Theorem 4.6. For the upper bounds see [90, 91] and [352], Chapter 4, §5, Remark 2. Finally, in the case $q = \infty$ the upper bounds can be easily derived from (5.16). The lower bounds follow from a general statement (see [340], Chapter 2, Theorem 5.7):

Proposition 5.8. *Let U_n be a bounded linear operator from \mathbf{H}_p^r to $\mathcal{T}(Q_n)$. Then for $1 \leq p \leq \infty$*

$$\sup_{f \in \mathbf{H}_p^r} \|f - U_n(f)\|_\infty \geq C(d, p) 2^{-(r-1/p)n} n^{d-1}.$$

Theorem 5.7 and Proposition 5.8 show that the sequence of operators T_n is optimal (in the sense of order) among all linear operators of approximating by means of the hyperbolic cross polynomials in $\mathcal{T}(Q_{n+d})$.

Finally, we study the situation $\theta < \infty$. We will see that the third index θ influences the estimates significantly. We refer to [90, 91] and the recent papers [298, 391, 300, 47, 99, 109, 371].

Theorem 5.9. *Let $1 \leq p, q, \theta \leq \infty$ and $r > 1/p$. Then we have the following relations*

(i) *For $p \geq q$,*

$$\sup_{f \in \mathbf{B}_{p,\theta}^r} \|f - T_n(f)\|_q \asymp 2^{-rn} n^{(d-1)(1-1/\theta)}.$$

(ii) For $p < q$,

$$\sup_{f \in \mathbf{B}_{p,\theta}^r} \|f - T_n(f)\|_q \asymp \begin{cases} 2^{-(r-1/p+1/q)n} n^{(d-1)(1/q-1/\theta)+}, & q < \infty, \\ 2^{-(r-1/p)n} n^{(d-1)(1-1/\theta)}, & q = \infty. \end{cases}$$

For the respective lower bounds we refer to [109, Sect. 5] and [371], see also Section 8.3 below. Note, that lower bounds for numerical integration also serve as lower bounds for approximation. Regarding upper bounds we have several options how to proceed. Starting with

$$\|f - T_n(f)\|_q = \left\| \sum_{|\mathbf{j}|_1 > n} v_{\mathbf{j}}(f) \right\|_q \quad (5.30)$$

one option is to use embeddings of Besov-Sobolev classes into L_q , see Lemma 3.12, in order to replace the right-hand side of (5.30) by an expression of type (5.22) or (5.23). Here we need Proposition 5.4 above combined with the embedding relations from Lemma 3.12. Another option is to simply use the triangle inequality for the error (5.30) and afterwards the Bernstein-Nikol'skii inequality (Theorem 2.6) to change the L_q -norm to an L_p -norm for instance. Then we get immediately an expression of type (5.22) or (5.23). In the latter expressions r will be $1/p - 1/q$ in the case $p < q$ (not the r from the class above). Using Hölder's inequality and standard procedures we end up with expressions like on the left-hand side of (5.18), (5.19) multiplied with a certain rate which is always generated from infinite sums of type

$$\left(\sum_{|\mathbf{j}|_1 > n} 2^{-|\mathbf{j}|_1 s \eta} \right)^{1/\eta} \asymp 2^{-sn} n^{(d-1)/\eta}. \quad (5.31)$$

Finally we apply our Proposition 5.3.

A general linear sampling operator on Smolyak grids $SG^d(n)$ is given by

$$\Psi(f, SG^d(n)) = \sum_{\mathbf{y} \in SG^d(n)} f(\mathbf{y}) \psi_{\mathbf{y}}, \quad (5.32)$$

where $\Psi_n = \{\psi_{\mathbf{y}}\}_{\mathbf{y} \in SG^d(n)}$ denotes a family of functions indexed by the grid points in $SG^d(n)$, see (8.4) below. On the basis of the time-limited B-spline sampling representation (5.44) which is given in Subsection 5.5 below, we construct the sampling algorithms R_n on the Smolyak grids $SG^d(n)$ by

$$R_n(f) := \sum_{\substack{\mathbf{j} \in \mathbb{N}_0^d \\ |\mathbf{j}|_1 \leq n}} q_{\mathbf{j}}(f) \quad , \quad f \in C(\mathbb{T}^d), \quad (5.33)$$

which induces a linear sampling algorithm of the form (5.32) where $\psi_{\mathbf{y}}$ are explicitly constructed as linear combinations of at most n_0 B-splines $N_{\mathbf{s},\mathbf{k}}$ for some $n_0 \in \mathbb{N}$ which is independent of $\mathbf{s}, \mathbf{k}, n$ and f . This fact can be proven in the same way as the proof of its counterpart for non-periodic functions [99]. Differing from T_n , the operator R_n does not possess an interpolation property similar to (5.25), which is why it is called *quasi-interpolation*.

Theorem 5.10. *All Theorems 5.5 – 5.9 hold true with frequency-limited sampling operators T_n replaced by the time-limited sampling operators R_n (with proper restrictions on r according to the smoothness of the B-spline).*

Theorem 5.10 is proved in [49], [103] for Sobolev classes \mathbf{W}_p^r . The upper bounds in Theorem 5.10 have been proven in [99, Theorem 3.1] for Besov classes $\mathbf{B}_{p,\theta}^r$ and the lower bounds follow from [109, Theorem 5.1].

Both sampling operators T_n and R_n provide approximate recovery with similar error bounds for classes \mathbf{W}_p^r and $\mathbf{B}_{p,\theta}^r$. The common feature of the operators T_n and R_n is that they use the Smolyak grid points as the sampling points. This motivates us to ask the following question. Are there (non-)linear operators, that use the Smolyak grids points for sampling, which give better bounds than T_n and R_n ? It is proved in [103, 109] that the answer to this question is “No”. We refer to a more general result in [109, Sect. 5] and [371], to establish the lower bounds in the theorems above. Note also that in case $p \geq q \geq 1$ the lower bounds for numerical integration also serve as lower bounds for approximation, see also Subsection 8.3 below. In general, the proof of the corresponding lower bounds is based on the construction of test functions from a class of our interest which is zero at the grid points. A nontrivial part of it is the proof of the fact that the constructed function belongs to the class. The inverse time-limited B-spline representation theorems (see [103, 109]) and Proposition 5.25, see [388], is used at this step.

Moreover, it is proved in [371] that even if we allow more general than the Smolyak grids sampling sets we do not gain better error bounds. We give a precise formulation of this result. Let $\mathbf{s} = (s_1, \dots, s_d)$, $s_j \in \mathbb{N}_0$, $j = 1, \dots, d$. We associate with \mathbf{s} a web $W(\mathbf{s})$ as follows: denote

$$w(\mathbf{s}, \mathbf{x}) := \prod_{j=1}^d \sin(2^{s_j} x_j)$$

and define

$$W(\mathbf{s}) := \{\mathbf{x} : w(\mathbf{s}, \mathbf{x}) = 0\}.$$

We say that a set of nodes $X_m := \{\mathbf{x}^i\}_{i=1}^m$ is an (n, ℓ) -net if $|X_m \setminus W(\mathbf{s})| \leq 2^\ell$ for all \mathbf{s} such that $|\mathbf{s}|_1 = n$. It is easy to check that $SG^d(n) \subset W(\mathbf{s})$ with any \mathbf{s} such that $|\mathbf{s}|_1 = n$. This means that $SG^d(n)$ is an (n, ℓ) -net for any ℓ .

Theorem 5.11. *For any recovering operator $\Psi(\cdot, X_m)$ with respect to a $(n, n-1)$ -net X_m we have for $1 \leq p < q < \infty$, $r > \beta$,*

$$\Psi(\mathbf{B}_{p,\theta}^r, X_m)_q \gtrsim 2^{-n(r-\beta)} n^{(d-1)(1/q-1/\theta)}, \quad \beta := 1/p - 1/q.$$

The reader may find some further results on Smolyak type algorithms in papers [320] and [321].

The exact recovery of trigonometric polynomials with frequencies in hyperbolic crosses from a discrete set of samples also plays a role in many applications. Several authors, see for instance [160], [122], [112] and the references therein, considered the problem of adapting the well-known fast Fourier transform to the sparse grid spacial discretization (HCFFT) for the recovery of multivariate trigonometric polynomials with frequencies on a hyperbolic cross, see Figure 8. However, there are some stability issues as [188] shows. This is related to the fact that the grid size still grows exponentially in d , however considerably slower than the full grid. In [189] the authors proposed to use discretization points generated by a oversampled lattice rule coming from numerical integration (see also Subsection 9.8). Due to the lattice structure one may use the classical one dimensional FFT here.

5.4 Sampling widths

Let us transfer the preceding results to the language of sampling widths $\varrho_m(\mathbf{F}, X)$, see (5.1). Based on the results in the previous subsection we can give reasonable upper bounds in those

situations. Note that, although we provided lower bounds for the Smolyak worst case sampling error, we do not have sharp lower bounds for the sampling numbers $\varrho_m(\mathbf{F}, X)$ in most of the situations. Due to the specific framework (linear sampling numbers) we can of course use linear widths for the estimates from below. This leads to sharp results in some of the cases, see also Section 9.7 for results where $p, q, \theta < 1$. Note that the upper bounds in Theorems 5.7, 5.9, 5.5 and 5.6 can be directly transferred to upper estimates for sampling numbers ϱ_m by taking the number

$$m := \#\widetilde{SG}^d(n) \asymp 2^n n^{d-1}$$

of grid points in a Smolyak grid of level n into account. Note, that this follows from (5.27).

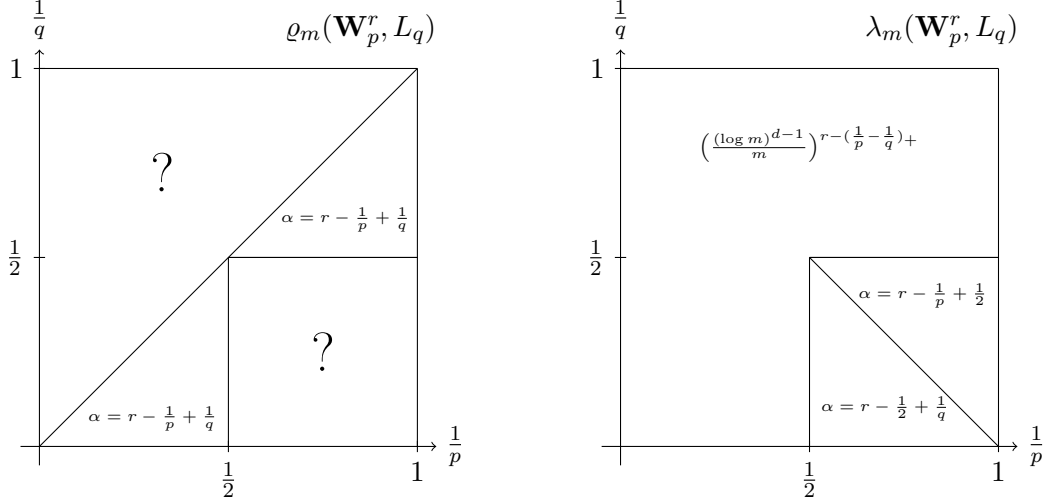


Figure 9: Comparison of $\varrho_m(\mathbf{W}_p^r, L_q)$ and $\lambda_m(\mathbf{W}_p^r, L_q)$, rate $(m^{-1} \log^{d-1} m)^\alpha$

Theorem 5.12. (i) Let $1 < p < q \leq 2$ and $r > 1/p$. Then

$$\varrho_m(\mathbf{W}_p^r, L_q) \asymp m^{-(r-1/p+1/q)} (\log m)^{(d-1)(r-1/p+1/q)}.$$

(ii) Let $2 \leq p < q < \infty$ and $r > 1/p$. Then

$$\varrho_m(\mathbf{W}_p^r, L_q) \asymp m^{-(r-1/p+1/q)} (\log m)^{(d-1)(r-1/p+1/q)}.$$

(iii) Let $r > 1/2$. Then we have

$$\varrho_m(\mathbf{W}_2^r, L_\infty) \asymp m^{-(r-1/2)} (\log m)^{(d-1)r}.$$

The results in (i) and (ii) have been proved recently in [48] for $r > 1/p$, and in [103] for $r > \max\{1/p, 1/2\}$. The special case $p = 2 < q$ is proved in [47, Thm. 6.10]. The result (iii) is proved in [351]. Clearly, (i) and (ii) follow from Subsection 4.5 together with Theorem 5.5. The upper bound in (iii) follows from Theorem 5.6. The lower bound is based on Ismagilov's [181] duality theorem

$$\lambda_m(\mathbf{W}_2^r, L_\infty) \asymp \lambda_m(\mathbf{W}_1^r, L_2)$$

(see also [59]) together with Theorem 4.17. This together with Theorem 4.13 also shows a sharp lower bound.

For completeness we reformulate (5.6) as follows.

Theorem 5.13. *Let $1 < p < q \leq 2$ and $r > 1/p$. Then we have*

$$\varrho_m(\mathbf{H}_p^r, L_q) \asymp m^{-(r-1/p+1/q)} (\log m)^{(d-1)(r-1/p+2/q)}.$$

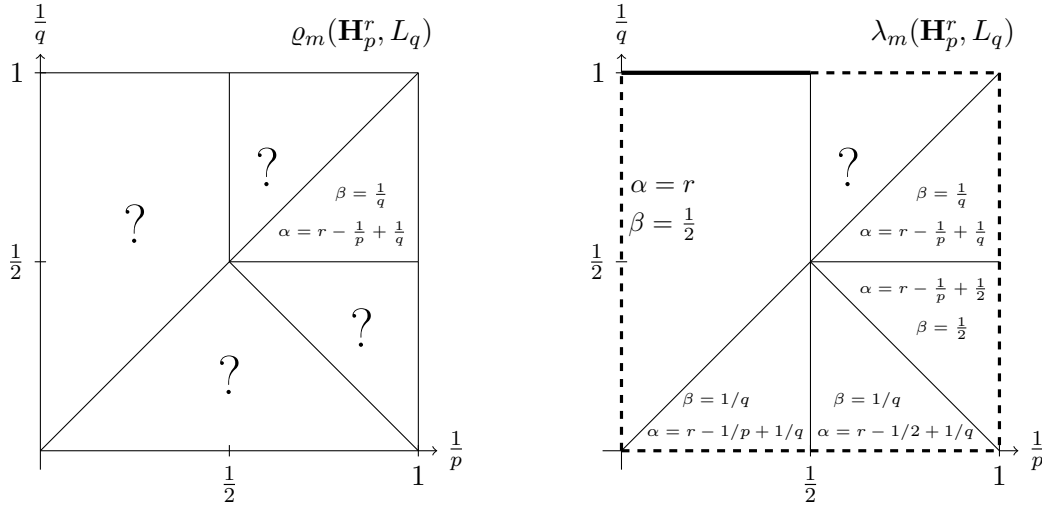


Figure 10: Comparison of $\varrho_m(\mathbf{H}_p^r, L_q)$ and $\lambda_m(\mathbf{H}_p^r, L_q)$

In the figure α and β refer to the asymptotic order

$$\left(\frac{(\log m)^{d-1}}{m} \right)^\alpha (\log m)^{(d-1)\beta}.$$

Remark 5.14. *It is pointed out in [371] that the upper bound in (5.3) and the lower bound for the Kolmogorov widths in Theorem 4.26, proved in [356], imply for $d = 2$, $r > 1/2$,*

$$\varrho_m(\mathbf{H}_\infty^r, L_\infty) \asymp m^{-r} (\log m)^{r+1}.$$

For $d > 2$ the correct order is not known.

Theorem 5.15. *Let $1 < p, q, \theta \leq \infty$ and $r > 1/p$. Then we have the following.*

(i) *For $p \geq q$,*

$$\varrho_m(\mathbf{B}_{p,1}^r, L_q) \asymp (m^{-1} \log^{d-1} m)^r, \quad \begin{cases} 2 \leq q < p < \infty, \\ 1 < p = q \leq \infty. \end{cases}$$

(ii) *For $1 < p < q < \infty$,*

$$\varrho_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp (m^{-1} \log^{d-1} m)^{r-1/p+1/q} (\log^{d-1} m)^{(1/q-1/\theta)_+}, \quad \begin{cases} 2 \leq p, 2 \leq \theta \leq q, \\ q \leq 2. \end{cases}$$

Proof. This theorem directly follows from Theorem 5.9 together with the lower bounds for linear widths in Theorem 4.47. Notice that it is also easily obtained from a non-periodic version in [99]. ■

5.5 Time-limited sampling representations–B-splines

This section is devoted to a second method of constructing sampling operators based on Smolyak’s algorithm. This time the approximant is not longer a trigonometric polynomial (like in (5.24)), it is rather a superposition of tensor products of compactly supported functions such as hat functions and more general B-splines. The potential of this technique for the approximation and integration of functions with dominating mixed smoothness has been recently observed by Triebel [382, 383] and, independently, Dinh Dũng [99]. The latter reference deals with B-spline representation, where the Faber-Schauder system is a special case. One striking advantage of this approach is its potential for non-periodic spaces of dominating mixed smoothness, see [99]. Apart from that the relations (5.40), (5.42), and (5.48) below are very well suited for construction “fooling functions” for sampling recovery and numerical integration, see Section 8.2 and Remark 8.9 below. In this subsection, for convenience we will use \mathbb{T} for the interval $[0, 1]$ (instead of $[0, 2\pi]$) with the usual identification of the end points.

The Faber-Schauder basis

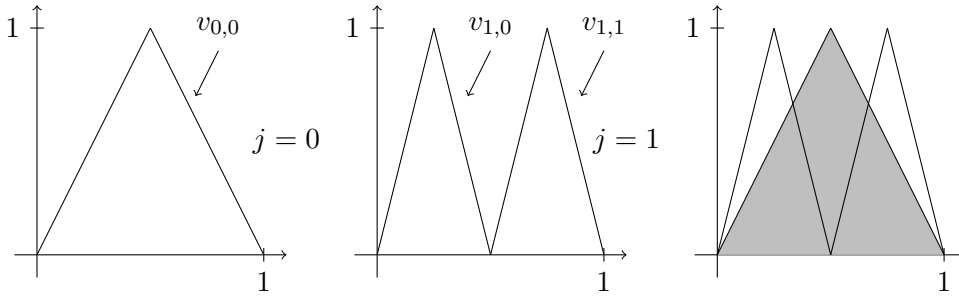


Figure 11: The univariate Faber-Schauder basis, levels $j = 0, 1$

Let us briefly recall the basic facts about the Faber-Schauder basis taken from [382, 3.2.1, 3.2.2]. Faber [120] observed that every continuous (non-periodic) function f on $[0, 1]$ can be represented (point-wise) as

$$f(x) = f(0) \cdot (1 - x) + f(1) \cdot x - \frac{1}{2} \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \Delta_{2^{-j-1}}^2(f, 2^{-j}k) v_{j,k}(x) \quad (5.34)$$

with convergence at least point-wise. Consequently, every periodic function on $C(\mathbb{T})$ can be represented by

$$f(x) = f(0) - \frac{1}{2} \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \Delta_{2^{-j-1}}^2(f, 2^{-j}k) v_{j,k}(x). \quad (5.35)$$

Definition 5.16. *The univariate periodic Faber-Schauder system is given by the system of functions on \mathbb{T}*

$$\{1, v_{j,k} : j \in \mathbb{N}_0, k \in \mathcal{D}_j\},$$

where $\mathcal{D}_j := \{0, \dots, 2^j - 1\}$ if $j \in \mathbb{N}_0$, $\mathcal{D}_{-1} := \{0\}$ and

$$v_{j,m}(x) = \begin{cases} 2^{j+1}(x - 2^{-j}m) & : 2^{-j}m \leq x \leq 2^{-j}m + 2^{-j-1}, \\ 2^{j+1}(2^{-j}(m+1) - x) & : 2^{-j}m + 2^{-j-1} \leq x \leq 2^{-j}(m+1), \\ 0 & : \text{otherwise.} \end{cases} \quad (5.36)$$

For notational reasons we let $v_{-1,0} := 1$ and obtain the Faber-Schauder system

$$\mathcal{F} := \{v_{j,k} : j \in \mathbb{N}_{-1}, k \in \mathcal{D}_j\}.$$

We denote by

$$v := v_{0,0}$$

the Faber basis function on level zero.

The tensor Faber-Schauder system

Let now $f(x_1, \dots, x_d)$ be a d -variate function $f \in C(\mathbb{T}^d)$. By fixing all variables except x_i we obtain by $g(\cdot) = f(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_d)$ a univariate periodic continuous function. By applying (5.35) in every such component we obtain the point-wise representation

$$f(\mathbf{x}) = \sum_{\mathbf{s} \in \mathbb{N}_{-1}^d} \sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{s}}} d_{\mathbf{s},\mathbf{k}}^2(f) v_{\mathbf{s},\mathbf{k}}(\mathbf{x}) \quad , \quad \mathbf{x} \in \mathbb{T}^d, \quad (5.37)$$

where $\mathcal{D}_{\mathbf{s}} = \mathcal{D}_{s_1} \times \dots \times \mathcal{D}_{s_d}$,

$$v_{\mathbf{s},\mathbf{k}}(x_1, \dots, x_d) := v_{s_1,k_1}(x_1) \cdot \dots \cdot v_{s_d,k_d}(x_d) \quad , \quad \mathbf{s} \in \mathbb{N}_{-1}^d, \mathbf{k} \in \mathcal{D}_{\mathbf{s}},$$

and

$$d_{\mathbf{s},\mathbf{k}}^2(f) := (-2)^{-|e(\mathbf{s})|} \Delta_{2^{-(\mathbf{s}+1)}}^{2,e(\mathbf{s})}(f, \mathbf{x}_{\mathbf{s},\mathbf{k}}) \quad , \quad \mathbf{s} \in \mathbb{N}_{-1}^d, \mathbf{k} \in \mathcal{D}_{\mathbf{s}}. \quad (5.38)$$

Here we put $e(\mathbf{s}) = \{i : s_i \neq -1\}$ and $\mathbf{x}_{\mathbf{s},\mathbf{k}} = (2^{-(j_1)+k_1}, \dots, 2^{-(j_d)+k_d})$.

The Faber-Schauder basis for Besov spaces

Our next goal is to discretize the spaces $\mathbf{B}_{p,\theta}^r$ using the Faber-Schauder system $\mathcal{F}^d := \{v_{\mathbf{s},\mathbf{k}} : \mathbf{s} \in \mathbb{N}_{-1}^d, \mathbf{k} \in \mathcal{D}_{\mathbf{s}}\}$. We obtain a sequence space isomorphism performed by the coefficient mapping $d_{\mathbf{s},\mathbf{k}}^2(f)$ above. In [382, 3.2.3, 3.2.4] and [99, Thm. 4.1] this was done for the non-periodic setting $\mathbf{B}_{p,\theta}^r(Q_d)$. For the results stated below we refer to the recent paper [175].

Definition 5.17. Let $0 < p, \theta \leq \infty$ and $r \in \mathbb{R}$. Then $\mathbf{b}_{p,\theta}^r$ is the collection of all sequences $\{\lambda_{\mathbf{s},\mathbf{k}}\}_{\mathbf{s} \in \mathbb{N}_{-1}^d, \mathbf{k} \in \mathcal{D}_{\mathbf{s}}}$ such that

$$\|\lambda_{\mathbf{s},\mathbf{k}}\|_{\mathbf{b}_{p,\theta}^r} := \left[\sum_{\mathbf{s} \in \mathbb{N}_{-1}^d} 2^{|\mathbf{s}|_1(r-1/p)q} \left(\sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{s}}} |\lambda_{\mathbf{s},\mathbf{k}}|^p \right)^{q/p} \right]^{1/q}$$

is finite.

Proposition 5.18. Let $1 \leq p, \theta \leq \infty$ and $1/p < r < 2$. Then there exists a constant $c > 0$ such that

$$\|d_{\mathbf{s},\mathbf{k}}^2(f)\|_{\mathbf{b}_{p,\theta}^r} \leq c \|f\|_{\mathbf{B}_{p,\theta}^r} \quad (5.39)$$

for all $f \in C(\mathbb{T}^d)$.

Let us also give a converse statement. Note, that we do not need the condition $r > 1/p$ here.

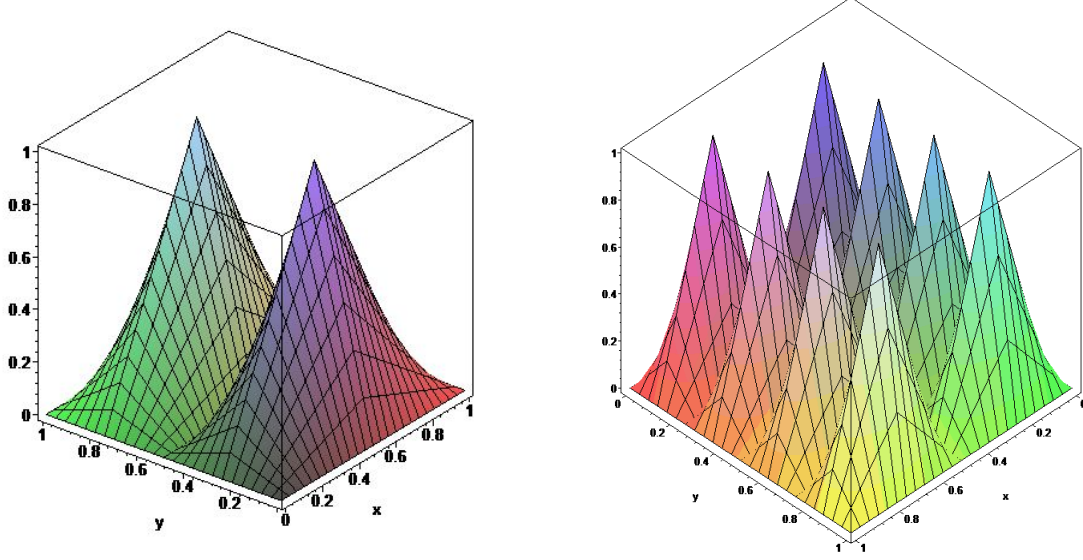


Figure 12: Faber-Schauder levels in $d = 2$

Proposition 5.19. *Let $1 \leq p, \theta \leq \infty$ and $0 < r < 1 + 1/p$. Let further $\{\lambda_{\mathbf{s}, \mathbf{k}}\}$ be a sequence belonging to $\mathbf{b}_{p, \theta}^r$. Then the function*

$$f := \sum_{\mathbf{s} \in \mathbb{N}_{-1}^d} \sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{s}}} \lambda_{\mathbf{s}, \mathbf{k}} v_{\mathbf{s}, \mathbf{k}}$$

belongs to $\mathbf{B}_{p, \theta}^r$ and

$$\|f\|_{\mathbf{B}_{p, \theta}^r} \leq c \|\lambda_{\mathbf{s}, \mathbf{k}}(f)\|_{\mathbf{b}_{p, \theta}^r}. \quad (5.40)$$

The Faber-Schauder basis for Sobolev spaces

Let us now come to the Sobolev spaces of mixed smoothness. We will state counterparts of the relations in Part (ii) of the Propositions 5.3, 5.4 above. Partial results have been already obtained in [382] if $r = 1$.

Definition 5.20. *Let $1 < p < \infty$ and $r \in \mathbb{R}$. Then \mathbf{w}_p^r is the collection of all sequences $\{\lambda_{\mathbf{s}, \mathbf{k}}\}_{\mathbf{s} \in \mathbb{N}_{-1}^d, \mathbf{k} \in \mathcal{D}_{\mathbf{s}}}$ such that*

$$\|\lambda_{\mathbf{s}, \mathbf{k}}\|_{\mathbf{w}_p^r} := \left\| \left[\sum_{\mathbf{s} \in \mathbb{N}_{-1}^d} 2^{|\mathbf{s}|_1 r/2} \left(\sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{s}}} |\lambda_{\mathbf{s}, \mathbf{k}}(f) v_{\mathbf{s}, \mathbf{k}}|^2 \right) \right]^{1/2} \right\|_p$$

is finite.

Proposition 5.21. *Let $1 < p < \infty$ and $\max\{1/p, 1/2\} < r < 2$ then we have for any $f \in C(\mathbb{T}^d)$*

$$\|d_{\mathbf{s}, \mathbf{k}}^2(f)\|_{\mathbf{w}_p^r} \lesssim \|f\|_{\mathbf{w}_p^r}. \quad (5.41)$$

Proposition 5.22. *Let $1 < p < \infty$ and $0 < r < \min\{1 + 1/p, 3/2\}$. Let furthermore $\lambda_{\mathbf{s}, \mathbf{k}} \in \mathbf{w}_p^r$. Then $f = \sum_{\mathbf{s} \in \mathbb{N}_{-1}^d} \sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{s}}} \lambda_{\mathbf{s}, \mathbf{k}} v_{\mathbf{s}, \mathbf{k}}$ belongs to \mathbf{W}_p^r and*

$$\|f\|_{\mathbf{W}_p^r} \lesssim \|\lambda_{\mathbf{s}, \mathbf{k}}\|_{\mathbf{w}_p^r}. \quad (5.42)$$

In case $r = 1$ both assertions can be found in [382, Chapt. 3]. In the stated form the relation will be rigorously proved in [49]. In fact, the result can be immediately deduced from its one dimensional counterpart via the tensor-product structure of the spaces and the corresponding operators, see [299, Thm. 2.1, 2.5]. Note, that Propositions 5.21, 5.22 imply that the tensorized Faber-Schauder system represents an unconditional basis in \mathbf{W}_p^r in the respective parameter domain. Let us illustrate this parameter domain in the following diagram.

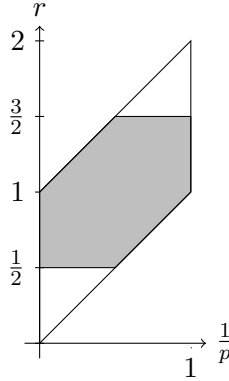


Figure 13: The parameter domain for the unconditional Faber-Schauder basis in \mathbf{W}_p^r

Remark 5.23. *There seems to be a fundamental difference between Besov spaces and Sobolev spaces in this context. A corresponding problem for Haar bases has been studied by Seeger, Ullrich in the recent papers [294, 295] which leads to the conclusion that the above described region is sharp. It also indicates that the parameter domain for the representation in Proposition 5.3, (ii) is sharp, i.e., it can not be extended to $1/p < r \leq 1/2$ if $p > 2$. This fundamental difference seems to be reflected in the unknown behavior of optimal cubature/sampling recovery in the region of small smoothness (lower triangles), see Section 8 below.*

B-spline representations and general atoms

When using the Faber-Schauder system for the discretization of function spaces we always get the restriction $r < 2$, see Propositions 5.18 and 5.19 above, due to the limited smoothness of the tensorized hat functions. The question arises whether one can use smoother basis functions such as B-splines. A smooth hierarchical basis approach similar to the one we will describe below has been developed by Bungartz [43]. Here we will focus on D. Dũng's [101, 103] approach towards a B-spline quasi-interpolation representation for periodic continuous functions on \mathbb{T}^d . For a non-periodic counterpart, see [99]. For a given $\ell \in \mathbb{N}$, denote by $M = M_{2\ell}$ the cardinal B-spline of order 2ℓ defined as the 2ℓ -fold convolution $M := (\chi_{[0,1]} * \cdots * \chi_{[0,1]})$, where $\chi_{[0,1]}$ denotes the characteristic function of the interval $[0, 1]$.

Let $\Lambda = \{\lambda(j)\}_{|j| \leq \mu}$ be a given finite even sequence, i.e., $\lambda(-j) = \lambda(j)$ for some $\mu \geq \ell - 1$. We define the linear operator Q for functions f on \mathbb{R} by

$$Q(f, x) := \sum_{k \in \mathbb{Z}} \Lambda(f, k) M(x - k), \quad \Lambda(f, k) := \sum_{|j| \leq \mu} \lambda(j) f(k - j + \ell). \quad (5.43)$$

The operator Q is called a *quasi-interpolation operator* if $Q(g) = g$ for every polynomial g of degree at most $2\ell - 1$. Since $M(2\ell 2^s x) = 0$ for every $s \in \mathbb{N}_0$ and every $x \notin (0, 1)$, we can extend the univariate B-spline $M(2\ell 2^s \cdot)$ to an 1-periodic function on the whole \mathbb{R} which

can be considered as a function on \mathbb{T} . Denote this function on \mathbb{T} by N_s and define $N_{s,k}(x) := N_s(x - (2\ell)^{-1}2^{-s}k)$, $s \in \mathbb{N}_0$, $k \in I(s)$, where $I(s) := \{0, 1, \dots, 2\ell 2^s - 1\}$. The quasi-interpolation operator Q induces the periodic quasi-interpolation operator Q_s on \mathbb{T} which is defined for $s \in \mathbb{N}_0$ and a function f on \mathbb{T} through

$$Q_s(f) := \sum_{k \in I(s)} a_{s,k}(f) N_{s,k} \quad , \quad a_{s,k}(f) := \sum_{|j| \leq \mu} \lambda(j) f((2\ell)^{-1}2^{-s}(k - j + \ell)).$$

A procedure similar to (5.15) yields an operator q_s such that every continuous function f on \mathbb{T}^d is represented as B -spline series

$$f = \sum_{\mathbf{s} \in \mathbb{N}_0^d} q_{\mathbf{s}}(f) = \sum_{\mathbf{s} \in \mathbb{N}_0^d} \sum_{\mathbf{k} \in I^d(\mathbf{s})} c_{\mathbf{s},\mathbf{k}}(f) N_{\mathbf{s},\mathbf{k}}, \quad (5.44)$$

converging in the norm of $C(\mathbb{T}^d)$, where $I^d(\mathbf{s}) := \prod_{i=1}^d I(s_i)$ and the coefficient functionals $c_{\mathbf{s},\mathbf{k}}(f)$ are explicitly constructed as linear combinations of at most n_0 function values of f for some $n_0 \in \mathbb{N}$ which is independent of \mathbf{s}, \mathbf{k} and f . The following proposition represents a counterpart of Proposition 5.3 and a generalization of Propositions 5.21 and 5.22.

Proposition 5.24. (i) *Let $1 \leq p, \theta \leq \infty$ and $1/p < r < \min\{2\ell, 2\ell - 1 + 1/p\}$. Then we have for any $f \in \mathbf{B}_{p,\theta}^r$,*

$$\left(\sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1 \theta} \|q_{\mathbf{j}}(f)\|_p^\theta \right)^{1/\theta} \asymp \|f\|_{\mathbf{B}_{p,\theta}^r} \quad (5.45)$$

with the sum being replaced by a supremum for $\theta = \infty$.

(ii) *Let $1 < p < \infty$ and $\max\{1/p, 1/2\} < r < 2\ell - 1$. Then we have for any $f \in \mathbf{W}_p^r$,*

$$\left\| \left(\sum_{\mathbf{j} \in \mathbb{N}_0^d} 2^{r|\mathbf{j}|_1^2} |q_{\mathbf{j}}(f)|^2 \right)^{1/2} \right\|_p \asymp \|f\|_{\mathbf{W}_p^r}. \quad (5.46)$$

Proposition 5.24 as well as a counterpart for the B -spline representation (5.44) of Proposition 5.4 have been proven in [103]. As in the proofs of Proposition 5.3(ii) and Propositions 5.21 and 5.22, the proof of Proposition 5.24(ii) requires tools from Fourier analysis, i.e., maximal functions of Peetre and Hardy-Littlewood type. Moreover, it is essentially based on a special explicit formula for the coefficients $c_{\mathbf{k},\mathbf{s}}(f)$ in the representation (5.44), and on the specific property of the representation (5.44) that the component functions $q_{\mathbf{s}}(f)$ can be split into a finite sum of the B -splines $N_{\mathbf{s},\mathbf{k}}$ having non-overlap interiors of their supports. One can probably extend the smoothness range in Proposition 5.24, (ii) to $\max\{1/p, 1/2\} < r < 2\ell - 1 + \min\{1/p, 1/2\}$ using similar techniques as in [49].

There are indeed many ways to construct quasi-interpolation operators built on B -splines, see, e.g., [58, 57, 39]. We give two examples of quasi-interpolation operators. For more examples, see [57]. A piecewise linear quasi-interpolation operator is defined as

$$Q(f, x) := \sum_{k \in \mathbb{Z}} f(k) M(x - k),$$

where M is the symmetric piecewise linear B -spline ($\ell = 1$). It is related to the classical Faber-Schauder basis of the hat functions (see, e.g., [99], [382], for details). Another example is the cubic quasi-interpolation operator generated by the symmetric cubic B -spline M ($\ell = 2$):

$$Q(f, x) := \sum_{k \in \mathbb{Z}} \frac{1}{6} \{-f(k-1) + 8f(k) - f(k+1)\} M(x - k).$$

We are interested in the most general form for C^∞ -bumps for all parameters $r > 0$. This has been shown by Vybíral [396] based on the approach of Frazier, Jawerth [127]. In fact, consider a smooth bump function φ (atom) supported in $[0, 1]^d$. We define

$$a_{\mathbf{s}, \mathbf{k}}(x_1, \dots, x_d) := \varphi(2^{s_1}x_1 - k_1) \cdot \dots \cdot \varphi(2^{s_d}x_d - k_d) \quad , \quad \mathbf{x} \in \mathbb{R}^d, \mathbf{s} \in \mathbb{N}_0^d, \mathbf{k} \in \mathbb{Z}^d.$$

Proposition 5.25. *Let $1 \leq p, \theta \leq \infty$ and $r > 0$. Let further $\{\lambda_{\mathbf{s}, \mathbf{k}}\}_{\mathbf{s}, \mathbf{k}}$ be a sequence belonging to $\mathbf{b}_{p, \theta}^r$. Then the function*

$$f := \sum_{\mathbf{s} \in \mathbb{N}_0^d} \sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{s}}} \lambda_{\mathbf{s}, \mathbf{k}} a_{\mathbf{s}, \mathbf{k}}(x) \quad (5.47)$$

exists and is supported in $[0, 1]^d$. Moreover,

$$\|f\|_{\mathbf{B}_{p, \theta}^r} \lesssim \left(\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{|\mathbf{s}|_1(r-1/p)\theta} \left[\sum_{\mathbf{k} \in \mathbb{Z}^d} |\lambda_{\mathbf{s}, \mathbf{k}}|^p \right]^{\theta/p} \right)^{1/\theta}. \quad (5.48)$$

5.6 Open problems

Below the reader may find a list of important open problems in this field. Afterwards we will comment on some of those.

Open problem 5.1 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{W}_p^r, L_p)$ in case $1 \leq p \leq \infty$ and $r > 1/p$.

Open problem 5.2 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{W}_p^r, L_q)$, $1 < p < 2 < q < \infty$, see also Figure 9.

Open problem 5.3 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{H}_p^r, L_q)$ in the question-marked regions in Figure 10 (left picture).

Open problem 5.4 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{W}_p^r, L_q)$ in the case of small smoothness, $2 < p < \infty$, $1/p < r \leq 1/2$.

Open problem 5.5 Find the right order of the optimal sampling recovery $\varrho_m(\mathbf{B}_{p, \theta}^r, L_p)$ for $\theta > 1$ and $\varrho_m(\mathbf{B}_{1, 1}^r, L_1)$. The latter problem seems to reduce to the problem of finding a lower bound for $\lambda_m(\mathbf{B}_{1, 1}^r, L_1)$, see Section 4.

The following simple example refers to Open problem 5.2 (see also the lower right triangle in Figure 9). It shows that even in case $d = 1$ sampling numbers ϱ_m and linear widths λ_m do not coincide.

Theorem 5.26. *Let $1 < p < 2 < q < \infty$ and $r > 1/p$. Then we have*

$$\begin{aligned} \lambda_m(\mathbf{W}_p^r, L_q) &\asymp m^{-(r-1/2+1/q)} (\log m)^{(d-1)(r-1/2+1/q)} \\ &= o(m^{-(r-1/p+1/q)}) \\ &= o(\varrho_m(\mathbf{W}_p^r, L_q)). \end{aligned} \quad (5.49)$$

Proof. Already in the univariate case it holds

$$m^{-(r-1/p+1/q)} \lesssim \varrho_m(\mathbf{W}_p^r, L_q).$$

■

The case $p = q$ looks rather simple. However, Open Problem 5.1 seems to be a hard problem. We state the following conjecture.

Conjecture 5.27. *Let $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$. Then*

$$\varrho_m(\mathbf{W}_p^r, L_p) \asymp m^{-r} (\log m)^{(d-1)(r+1/2)}.$$

The upper bound is known, see Theorem 5.5(i). Problematic is the lower bound. The linear widths are smaller than the bound in Conjecture 5.27, see Theorem 4.45. In other words, we conjecture two things. First, sampling is worse than approximation also in this situation and second, Smolyak's algorithm is optimal for sampling numbers. Note, that the Hilbert space situation \mathbf{W}_2^r in L_2 is also open in this respect.

Let us comment on Open problem 5.4. In the case of small smoothness we consider the situation $2 < p < \infty$ and $1/p < r \leq 1/2$. This problem is also relevant for numerical integration, see Section 8 below, where we were able to give a partial answer. In fact, it is possible to prove an interesting upper bound for $\kappa_m(\mathbf{W}_p^r)$ in the situation $2 < p < \infty$ and $1/p < r \leq 1/2$.

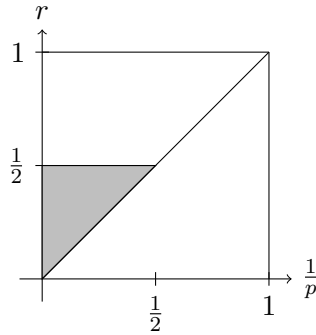


Figure 14: The region of “small smoothness”

Let us start with an interesting observation for small r in the Besov setting. Here the “difference” between λ_m and ϱ_m will be much smaller than in Theorem 5.26 and only apparent if $d \geq 2$.

Theorem 5.28. *Let $2 < p \leq \infty$ and $1/p < r < 1/2$. Then*

$$\begin{aligned} \lambda_m(\mathbf{B}_{p,p}^r, L_p) &\asymp m^{-r} (\log m)^{(d-1)(r+1/2-1/p)} \\ &= o(m^{-r} (\log m)^{(d-1)(1-1/p)}) \\ &= o(\varrho_m(\mathbf{B}_{p,p}^r, L_p)). \end{aligned} \tag{5.50}$$

Proof. Note that

$$m^{-r} (\log m)^{(d-1)(1-1/p)} \asymp \kappa_m(\mathbf{B}_{p,p}^r) \lesssim \varrho_m(\mathbf{B}_{p,p}^r, L_p) \lesssim m^{-r} (\log m)^{(d-1)(r+1-1/p)}.$$

■

6 Entropy numbers

6.1 General notions and inequalities

The concept of entropy is also known as *Kolmogorov entropy* and *metric entropy*. This concept allows us to measure how big is a compact set. In the case of finite dimensional compacts it is convenient to compare compact sets by their volumes. In the case of infinite dimensional Banach spaces this way does not work. The concept of entropy is a good replacement of the concept of volume in infinite dimensional Banach spaces.

Let X be a Banach space and let B_X denote the unit ball of X with the center at 0. Denote by $B_X(y, r)$ a ball with center y and radius r : $\{x \in X : \|x - y\| \leq r\}$. For a compact set A and a positive number ε we define the covering number $N_\varepsilon(A)$ as follows

$$N_\varepsilon(A) := N_\varepsilon(A, X) := \min\{n : \exists y^1, \dots, y^n : A \subseteq \cup_{j=1}^n B_X(y^j, \varepsilon)\}.$$

It is convenient to consider along with the entropy $H_\varepsilon(A, X) := \log_2 N_\varepsilon(A, X)$ the entropy numbers $\epsilon_n(A, X)$:

$$\epsilon_n(A, X) := \inf\{\varepsilon > 0 : \exists y^1, \dots, y^{2^n} \in X : A \subseteq \cup_{j=1}^{2^n} B_X(y^j, \varepsilon)\}.$$

The definition of the entropy numbers can be written in a form similar to the definition of the Kolmogorov widths (numbers):

$$\epsilon_n(A, X) := \inf_{y^1, \dots, y^{2^n}} \sup_{f \in A} \inf_j \|f - y^j\|_X.$$

In the case of finite dimensional spaces \mathbb{R}^d equipped with different norms, say, norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ the volume argument gives some bounds on the $N_\varepsilon(B_Y, X)$. For a Lebesgue measurable set $E \subset \mathbb{R}^d$ we denote its Lebesgue measure by $\text{vol}(E) := \text{vol}_d(E)$. Let further $C = A \oplus B := \{c : c = a + b, a \in A, b \in B\}$.

Theorem 6.1. *For any two norms X and Y and any $\varepsilon > 0$ we have*

$$\frac{1}{\varepsilon^d} \frac{\text{vol}(B_Y)}{\text{vol}(B_X)} \leq N_\varepsilon(B_Y, X) \leq \frac{\text{vol}(B_Y(0, 2/\varepsilon) \oplus B_X)}{\text{vol}(B_X)}.$$

Let us formulate one immediate corollary of Theorem 6.1.

Corollary 6.2. *For any d -dimensional Banach space X we have*

$$\varepsilon^{-d} \leq N_\varepsilon(B_X, X) \leq (1 + 2/\varepsilon)^d,$$

and, therefore,

$$\epsilon_n(B_X, X) \leq 3(2^{-n/d}).$$

Let us consider some typical d -dimensional Banach spaces. These are the spaces ℓ_p^d : the linear space \mathbb{R}^d equipped with the norms

$$|x|_p := \|x\|_{\ell_p^d} := \left(\sum_{j=1}^d |x_j|^p \right)^{1/p}, \quad 1 \leq p < \infty,$$

$$|x|_\infty := \|x\|_{\ell_\infty^d} := \max_j |x_j|.$$

Denote $B_p^d := B_{\ell_p^d}$.

Theorem 6.3. *Let $0 < p < q \leq \infty$. Then we have the following bounds for the entropy numbers $\epsilon_n(B_p^d, \ell_q^d)$*

$$\epsilon_n(B_p^d, \ell_q^d) \asymp \begin{cases} 1, & 1 \leq n \leq \log_2 d; \\ \left[\frac{\log(\frac{d}{n}+1)}{n} \right]^{\frac{1}{p}-\frac{1}{q}}, & \log_2 d \leq n \leq d; \\ d^{\frac{1}{q}-\frac{1}{p}} 2^{-n/d}, & d \leq n. \end{cases}$$

Theorem 6.3 has been obtained by Schütt [285] for the Banach-space range of parameters. He considered the more general situation of symmetric Banach spaces, see also [178] and [226]. For the quasi-Banach situation we refer to [119] and finally to Kühn [207], Guédon, Litvak [156] and Edmunds, Netrusov [117, Thm. 2], where the lower bound in the middle case is provided.

We define by $\sigma(x)$ the normalized $(d-1)$ -dimensional measure on the sphere S^{d-1} – the boundary of B_2^d .

Theorem 6.4. *Let X be \mathbb{R}^d equipped with $\|\cdot\|$ and*

$$M_X := \int_{S^{d-1}} \|x\| d\sigma(x).$$

Then we have

$$\epsilon_n(B_2^d, X) \lesssim M_X \begin{cases} (d/n)^{1/2}, & n \leq d \\ 2^{-n/d}, & n \geq d. \end{cases}$$

Theorem 6.4 is a dual version of the corresponding result from [312]. Theorem 6.4 was proved in [256].

There are several general results which give lower estimates of the Kolmogorov widths $d_n(F, X)$ in terms of the entropy numbers $\epsilon_k(F, X)$.

Theorem 6.5. *For any $r > 0$ we have*

$$\max_{1 \leq k \leq n} k^r \epsilon_k(F, X) \leq C(r) \max_{1 \leq m \leq n} m^r d_{m-1}(F, X). \quad (6.1)$$

Theorem 6.5 is due to Carl [51], which is actually stated for general s -numbers and unit balls F stemming from a norm, see also Carl, Stephani [53]. Recently, Hinrichs et al. [174] proved a version for quasi-Banach spaces (non-convex F), which is particularly interesting for Gelfand numbers c_m on the right-hand side, see Subsection 9.6 below.

Let us introduce a nonlinear Kolmogorov's (N, m) -width:

$$d_m(F, X, N) := \inf_{\Lambda_N, \#\Lambda_N \leq N} \sup_{f \in F} \inf_{L \in \Lambda_N} \inf_{g \in L} \|f - g\|_X,$$

where Λ_N is a set of at most N m -dimensional subspaces L . It is clear that

$$d_m(F, X, 1) = d_m(F, X).$$

The new feature of $d_m(F, X, N)$ is that we allow to choose a subspace $L \in \Lambda_N$ depending on $f \in F$. It is clear that the bigger N the more flexibility we have to approximate f . The following inequality from [358] (see also [368], Section 3.5) is a generalization of inequality (6.1).

Theorem 6.6. *Let $r > 0$. Then*

$$\max_{1 \leq k \leq n} k^r \epsilon_k(F, X) \leq C(r, K) \max_{1 \leq m \leq n} m^r d_{m-1}(F, X, K^m), \quad (6.2)$$

where we denote

$$d_0(F, X, N) := \sup_{f \in F} \|f\|_X.$$

The possibility of replacing K^m by $(Kn/m)^m$ in (6.2) was discussed in [358], in Section 3.5 of [368], and in [369]. The corresponding remark is from [369].

Remark 6.7. *Examining the proof of (6.2) one can check that the following inequality holds*

$$n^r \epsilon_n(F, X) \leq C(r, K) \max_{1 \leq m \leq n} m^r d_{m-1}(F, X, (Kn/m)^m).$$

Finally, Theorem 6.8 is from [369].

Theorem 6.8. *Let a compact $F \subset X$ and a number $r > 0$ be such that for some $n \in \mathbb{N}$*

$$d_{m-1}(F, X, (Kn/m)^m) \leq m^{-r}, \quad m \leq n.$$

Then for $k \leq n$

$$\epsilon_k(F, X) \leq C(r, K) \left(\frac{\log(2n/k)}{k} \right)^r.$$

We discuss an application which motivated a study of $d_m(F, X, N)$ with $N = (Kn/m)^m$. Let $\mathcal{D} = \{g_j\}_{j=1}^n$ be a system of normalized elements of cardinality $|\mathcal{D}| = n$ in a Banach space X . Consider best m -term approximations of f with respect to \mathcal{D}

$$\sigma_m(f, \mathcal{D})_X := \inf_{\{c_j\}; |\Lambda|=m} \left\| f - \sum_{j \in \Lambda} c_j g_j \right\|.$$

For a function class F set

$$\sigma_m(F, \mathcal{D})_X := \sup_{f \in F} \sigma_m(f, \mathcal{D})_X.$$

Then it is clear that for any system \mathcal{D} , $|\mathcal{D}| = n$,

$$d_m(F, X, \binom{n}{m}) \leq \sigma_m(F, \mathcal{D})_X.$$

Next,

$$\binom{n}{m} \leq (en/m)^m.$$

Thus Theorem 6.8 implies the following theorem (see [369]).

Theorem 6.9. *Let a compact $F \subset X$ be such that there exists a normalized system \mathcal{D} , $|\mathcal{D}| = n$, and a number $r > 0$ such that*

$$\sigma_m(F, \mathcal{D})_X \leq m^{-r}, \quad m \leq n.$$

Then for $k \leq n$

$$\epsilon_k(F, X) \leq C(r) \left(\frac{\log(2n/k)}{k} \right)^r.$$

Theorem 6.9 is useful in proving lower bounds for best m -term approximations. Recently, it was shown in [374] how Theorem 6.9 can be used for proving sharp upper bounds for the entropy numbers of classes of mixed smoothness.

We now proceed to two multiplicative inequalities for the L_p spaces. Let D be a domain in \mathbb{R}^d and let $L_p := L_p(D)$ denote the corresponding L_p space, $1 \leq p \leq \infty$, with respect to the Lebesgue measure. We note that the inequalities below hold for any measure μ on D .

Theorem 6.10. *Let $A \subset L_1 \cap L_\infty$. Then for any $1 \leq p \leq \infty$ we have $A \subset L_p$ and*

$$\epsilon_{n+m}(A, L_p) \leq 2\epsilon_n(A, L_1)^{1/p} \epsilon_m(A, L_\infty)^{1-1/p}.$$

It will be convenient for us to formulate one more inequality in terms of entropy numbers of operators. Let S be a linear operator from X to Y . We define the n th entropy number of S as

$$\epsilon_n(S : X \rightarrow Y) := \epsilon_n(S(B_X), Y)$$

where $S(B_X)$ is the image of B_X under mapping S .

Theorem 6.11. *For any $1 \leq p \leq \infty$ and any Banach space Y we have*

$$\epsilon_{n+m}(S : L_p \rightarrow Y) \leq 2\epsilon_n(S : L_1 \rightarrow Y)^{1/p} \epsilon_m(S : L_\infty \rightarrow Y)^{1-1/p}.$$

6.2 Entropy numbers for \mathbf{W} classes in L_q

It is well known that in the univariate case

$$\epsilon_n(W_p^r, L_q) \asymp n^{-r} \tag{6.3}$$

holds for all $1 \leq p, q \leq \infty$ and $r > (1/p - 1/q)_+$. We note that condition $r > (1/p - 1/q)_+$ is a necessary and sufficient condition for a compact embedding of W_p^r into L_q . Thus (6.3) provides a complete description of the rate of $\epsilon_n(W_p^r, L_q)$ in the univariate case. We point out that (6.3) shows that the rate of decay of $\epsilon_n(W_p^r, L_q)$ depends only on r and does not depend on p and q . In this sense the strongest upper bound (for $r > 1$) is $\epsilon_n(W_1^r, L_\infty) \lesssim n^{-r}$ and the strongest lower bound is $\epsilon_n(W_\infty^r, L_1) \gtrsim n^{-r}$.

There are different generalizations of classes W_p^r to the case of multivariate functions. In this section we only discuss classes \mathbf{W}_p^r of functions with bounded mixed derivative and $\mathbf{B}_{p,\theta}^r$ of functions with bounded mixed difference. These classes are of special interest for several reasons:

- (A) The problem of the rate of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$ in a particular case $r = 1, p = 2, q = \infty$ is equivalent (see [206]) to a fundamental problem of probability theory (the small ball problem). Both of these problems are still open for $d > 2$.
- (B) The problem of the rate of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$ and $\epsilon_n(\mathbf{B}_{p,\theta}^r, L_q)$ turns out to be a very rich and difficult problem. There are still many open problems. Those problems that have been resolved required different nontrivial methods for different pairs (p, q) .

We begin with classes \mathbf{W}_p^r .

Theorem 6.12. *For $1 < p, q < \infty$ and $r > (1/p - 1/q)_+$ one has*

$$\epsilon_n(\mathbf{W}_p^r, L_q) \asymp n^{-r} (\log n)^{r(d-1)}. \tag{6.4}$$

Theorem 6.13. *For $r > 0$ and $1 \leq p < \infty$ one has*

$$\epsilon_n(\mathbf{W}_p^r, L_1) \gtrsim n^{-r} (\log n)^{r(d-1)}.$$

The problem of estimating $\epsilon_n(\mathbf{W}_p^r, L_q)$ has a long history. The first result on the right order of $\epsilon_n(\mathbf{W}_2^r, L_2)$ was obtained by Smolyak [304]. The case $1 < q = p < \infty$, $r > 0$ was established by Dinh Dũng [87]. Theorem 6.12 was established by Temlyakov, see [342], [344] for all $1 < p, q < \infty$ and $r > 1$. Belinskii [34] extended (6.4) to the case $r > (1/p - 1/q)_+$. Later, this result was extended by Vybiral to a non-periodic setting, see [396].

It is known in approximation theory (see [352]) that investigation of asymptotic characteristics of classes \mathbf{W}_p^r in L_q becomes more difficult when p or q takes value 1 or ∞ than when $1 < p, q < \infty$. This is true for $\epsilon_n(\mathbf{W}_p^r, L_q)$, too. Theorem 6.13 is taken from [342, 344]. It was discovered that in some of these extreme cases (p or q equals 1 or ∞) relation (6.4) holds and in other cases it does not hold. We describe the picture in detail. It was proved in [344] that (6.4) holds for $q = 1$, $1 < p < \infty$, $r > 0$. It was also proved that (6.4) holds for $q = 1$, $p = \infty$ (see [34] for $r > 1/2$ and [196] for $r > 0$). Summarizing, we state that (6.4) holds for $1 < p, q < \infty$ and $q = 1$, $1 < p \leq \infty$ for all d (with appropriate restrictions on r). This easily implies that (6.4) also holds for $p = \infty$, $1 \leq q < \infty$. We formulate this as a theorem.

Theorem 6.14. *Let $q = 1$, $1 < p \leq \infty$ or $p = \infty$, $1 \leq q < \infty$ and $r > 0$. Then*

$$\epsilon_n(\mathbf{W}_p^r, L_q) \asymp n^{-r}(\log n)^{r(d-1)}.$$

For all other pairs (p, q) , namely, for $q = \infty$, $1 \leq p \leq \infty$ and $p = 1$, $1 \leq q \leq \infty$ the rate of $\epsilon_n(\mathbf{W}_p^r, L_q)$ is not known in the case $d > 2$. It is an outstanding open problem.

Entropy numbers in L_∞

Let us start with the following result for the case $d = 2$.

Theorem 6.15. *Let $d = 2$, $1 < p \leq \infty$, $r > \max\{1/p, 1/2\}$. Then*

$$\epsilon_n(\mathbf{W}_p^r, L_\infty) \asymp n^{-r}(\log n)^{r+1/2}. \quad (6.5)$$

The first result on the right order of $\epsilon_n(\mathbf{W}_p^r, L_q)$ in the case $q = \infty$ was obtained by Kuelbs and Li [206] for $p = 2$, $r = 1$. It was proved in [354] that (6.5) holds for $1 < p < \infty$, $r > 1$. We note that the upper bound in (6.5) was proved under condition $r > 1$ and the lower bound in (6.5) was proved under condition $r > 1/p$. Belinskii [34] proved the upper bound in (6.5) for $1 < p < \infty$ under condition $r > \max\{1/p, 1/2\}$. Relation (6.5) for $p = \infty$ under assumption $r > 1/2$ was proved in [357].

An analogue of the upper bound in (6.5) for any d was obtained by Belinskii [31, 34] and, independently, by Dunker, Linde, Lifshits, Kühn [116] (in case $r = 1$)

$$\epsilon_n(\mathbf{W}_p^r, L_\infty) \lesssim n^{-r}(\log n)^{(d-1)r+1/2}, \quad r > \max\{1/p, 1/2\}. \quad (6.6)$$

That proof is based on Theorem 6.4 (see also the book [385] for a detailed description of the technique). Recent results on the Small Ball Inequality for the Haar system (see [36], [37] and the remark after (2.29) in Subsection 2.6 above) allow us to improve a trivial lower bound to the following one for $r = 1$ and all $p < \infty$, $d \geq 3$:

$$\epsilon_n(\mathbf{W}_p^1, L_\infty) \gtrsim n^{-1}(\log n)^{d-1+\delta(d)}, \quad 0 < \delta(d) < 1/2.$$

Theorem 6.15 and the above upper and lower bounds support the following conjecture.

Conjecture 6.16. *Let $d \geq 3$, $2 \leq p \leq \infty$, $r > 1/2$. Then*

$$\epsilon_n(\mathbf{W}_p^r, L_\infty) \asymp n^{-r}(\log n)^{(d-1)r+1/2}.$$

It is known that the corresponding lower bound in Conjecture 6.16 would follow from the d -dimensional version of the Small Ball Inequality for the trigonometric system (2.28).

The case $p = 1$, $1 \leq q \leq \infty$ was settled by Kashin and Temlyakov [199]. The authors proved the following results.

Theorem 6.17. *Let $d = 2$, $1 \leq q < \infty$, $r > \max\{1/2, 1 - 1/q\}$. Then*

$$\epsilon_n(\mathbf{W}_1^r, L_q) \asymp n^{-r} (\log n)^{r+1/2} \quad (6.7)$$

Theorem 6.18. *Let $d = 2$, $r > 1$. Then*

$$\epsilon_n(\mathbf{W}_{1,0}^r, L_\infty) \asymp n^{-r} (\log n)^{r+1}. \quad (6.8)$$

The most difficult part of Theorems 6.17 and 6.18 – the lower bounds – is proved with the help of the volume estimates of the appropriate sets of the Fourier coefficients of bounded trigonometric polynomials. With the notation from Subsection 2.5 the volume estimates of the sets $B_\Lambda(L_p)$ and related questions have been studied in a number of papers: the case $\Lambda = [-n, n]$, $p = \infty$ in [193]; the case $\Lambda = [-N_1, N_1] \times [-N_2, N_2]$, $p = \infty$ in [344], [350]; the case of arbitrary Λ and $p = 1$ in [195]. In particular, the results of [195] imply for $d = 2$ and $1 \leq p < \infty$ that

$$(\text{vol}(B_{\Delta Q_n}(L_p)))^{(2|\Delta Q_n|)^{-1}} \asymp |\Delta Q_n|^{-1/2} \asymp (2^n n)^{-1/2}.$$

It was proved in [199] that in the case $p = \infty$ the volume estimate is different (see Theorem 2.23):

$$(\text{vol}(B_{\Delta Q_n}(L_\infty)))^{(2|\Delta Q_n|)^{-1}} \asymp (2^n n^2)^{-1/2}. \quad (6.9)$$

We note that in the case $\Lambda = [-N_1, N_1] \times [-N_2, N_2]$ the volume estimate is the same for all $1 \leq p \leq \infty$. The volume estimate (6.9) plays the key role in the proof of (6.7) and (6.8).

Let us make an observation on the base of the above discussion. In the univariate case the entropy numbers $\epsilon_n(W_p^r, L_q)$ have the same order of decay with respect to n for all pairs (p, q) , $1 \leq p, q \leq \infty$. In the case $d = 2$ we have three different orders of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$ which depend on the pair (p, q) . For instance, in the case $1 < p, q < \infty$ it is $n^{-r} (\log n)^r$, in the case $p = 1$, $1 < q < \infty$, it is $n^{-r} (\log n)^{r+1/2}$ and in the case $p = 1$, $q = \infty$ it is $n^{-r} (\log n)^{r+1}$.

We discussed above known results on the rate of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$. In the case $d = 2$ the picture is almost complete. In the case $d > 2$ the situation is fundamentally different. The problem of the right order of decay of $\epsilon_n(\mathbf{W}_p^r, L_q)$ is still open for $p = 1$, $1 \leq q \leq \infty$ and $q = \infty$, $1 \leq p \leq \infty$. In particular, it is open in the case $p = 2$, $q = \infty$, $r = 1$ that is related to the small ball problem. We discussed in more detail the case $q = \infty$, $1 \leq p \leq \infty$. We pointed out above that in the case $d = 2$ the proof of lower bounds (the most difficult part) was based on the Small Ball Inequalities for the Haar system for $r = 1$ and for the trigonometric system for all r .

6.3 Entropy numbers for \mathbf{H} and \mathbf{B} classes in L_q

Let us proceed with the entropy numbers of the \mathbf{H} classes in L_q .

Theorem 6.19. *For $r > 1$ one has*

$$\epsilon_n(\mathbf{H}_1^r, \mathbf{B}_{\infty,2}^0) \lesssim n^{-r} (\log n)^{(r+\frac{1}{2})(d-1)}.$$

Note that the space $\mathbf{B}_{\infty,2}^0$ is “close” to $L_\infty(\mathbb{T}^d)$, however not comparable. In fact, we have

$$\mathbf{B}_{\infty,1}^0 \hookrightarrow L_\infty(\mathbb{T}^d) \hookrightarrow \mathbf{B}_{\infty,\infty}^0.$$

Theorem 6.20. *For $r > 0$ one has*

$$\epsilon_n(\mathbf{H}_\infty^r, L_1) \gtrsim n^{-r} (\log n)^{(r+\frac{1}{2})(d-1)}.$$

Theorems 6.19 and 6.20 were obtained in [342] and [344]. Theorem 6.20 was proved earlier by N.S. Bakhvalov [18] in the following special cases: (I) $d = 2$; (II) any d but L_1 is replaced by L_2 . These theorems give the right order of $\epsilon_n(\mathbf{H}_p^r, L_q)$ for all $1 \leq p, q \leq \infty$, except the case $q = \infty$. In this case we have the following result by Belinskii [34].

Theorem 6.21. *Let $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$. Then*

$$\left(\frac{(\log n)^{d-1}}{n}\right)^r (\log n)^{(d-1)/2} \lesssim \epsilon_n(\mathbf{H}_p^r, L_\infty) \lesssim \left(\frac{(\log n)^{d-1}}{n}\right)^r (\log n)^{d/2}.$$

The following theorem shows that in the case $d = 2$ the upper bounds in Theorem 6.21 are sharp (in the sense of order).

Theorem 6.22. *In the case $d = 2$ for any $1 \leq p \leq \infty$, $r > 1/p$ we have*

$$\epsilon_n(\mathbf{H}_p^r, L_\infty) \asymp n^{-r} (\log n)^{r+1}.$$

One can prove the upper bounds in Theorem 6.22 using the discretization method based directly on discretization and on Theorem 6.3. The difficult part of Theorem 6.22 is the lower bound for $p = \infty$. This lower bound was proved in [354].

In the same way as we argued before formulating Conjecture 6.16 we can argue in support of the following conjecture.

Conjecture 6.23. *Let $d \geq 3$, $1 \leq p \leq \infty$, $r > 1/p$. Then*

$$\epsilon_n(\mathbf{H}_p^r, L_\infty) \asymp n^{-r} (\log n)^{(d-1)(r+1/2)+1/2}.$$

We discussed above results on the right order of decay of the entropy numbers. Clearly, each order relation \asymp is a combination of the upper bound \lesssim and the matching lower bound \gtrsim . We now briefly discuss methods that were used for proving upper and lower bounds. The upper bounds in Theorems 6.19 and 6.12 were proved by the standard method of reduction by discretization to estimates of the entropy numbers of finite-dimensional sets. Theorem 6.3 plays a key role in this method. It is clear from the above discussion that it was sufficient to prove the lower bound in (6.4) in the case $q = 1$. The proof of this lower bound (see Theorems 6.20 and 6.13) is more difficult and is based on nontrivial estimates of the volumes of the sets of Fourier coefficients of bounded trigonometric polynomials. Theorem 2.18 plays a key role in this method.

We continue with the $\mathbf{B}_{p,\theta}^r$ classes. A non-periodic version of the following result has been proved by Vybíral [396, Thms. 3.19, 4.11].

Theorem 6.24. *Let $0 < p, \theta \leq \infty$ and $1 < q < \infty$.*

(i) *If $r > (1/p - 1/q)_+$. Then*

$$\epsilon_n(\mathbf{B}_{p,\theta}^r, L_q) \gtrsim n^{-r} (\log n)^{(d-1)(r+1/2-1/\theta)_+}.$$

(ii) *If $r > (1/p - 1/q)_+ + 1/\min\{p, q, \theta\} - 1/\min\{p, q\} + 1/q - 1/\max\{q, 2\}$ then*

$$\epsilon_n(\mathbf{B}_{p,\theta}^r, L_q) \lesssim n^{-r} (\log n)^{(d-1)(r+1/2-1/\theta)}.$$

(iii) *If $r > \max\{1/p - 1/2, 1/\theta - 1/2\}$ then*

$$\epsilon_n(\mathbf{B}_{p,\theta}^r, L_2) \asymp n^{-r} (\log n)^{(d-1)(r+1/2-1/\theta)}.$$

Note, that the third index θ influences the power of the logarithm. There is an interesting effect of small smoothness observed by Vybíral in [396, Thm. 4.11]. The condition in Theorem 6.24, (iii) seems to be crucial in the following sense. If $\theta < p$ and $1/p - 1/2 < r \leq 1/\theta - 1/2$ then for any $\varepsilon > 0$ there is $c_\varepsilon > 0$ such that

$$\epsilon_n(\mathbf{B}_{p,\theta}^r, L_2) \leq c_\varepsilon n^{-r} (\log n)^\varepsilon. \quad (6.10)$$

The influence of the small mixed smoothness r seems to be dominated by the influence of the fine parameter θ . The correct order is not known, see also Open problem 6.4 below.

Note, that if $r > 1/\theta - 1/2$ the following stronger lower bound

$$\epsilon_n(\mathbf{B}_{\infty,\theta}^r, L_1) \gtrsim n^{-r} (\log n)^{(d-1)(r+1/2-1/\theta)}$$

follows directly from the proof of Theorem 6.20 in [344] and the inequality: for $f \in \mathcal{T}(\Delta Q_l)$

$$\|f\|_{\mathbf{B}_{p,\theta}^r} \lesssim l^{(d-1)/\theta} \|f\|_{\mathbf{H}_p^r}.$$

In the case $p \geq 2$ and $\theta \geq 2$ the corresponding upper bound for the $\epsilon_n(\mathbf{B}_{p,\theta}^r, L_q)$ can be derived from Theorem 6.12 with the help of inequalities

$$\left\| \sum_{|\mathbf{s}|_1=l} \delta_{\mathbf{s}}(f) \right\|_p \lesssim \left(\sum_{|\mathbf{s}|_1=l} \|\delta_{\mathbf{s}}(f)\|_p^2 \right)^{1/2} \lesssim \left(\sum_{|\mathbf{s}|_1=l} \|\delta_{\mathbf{s}}(f)\|_p^\theta \right)^{1/\theta} l^{(d-1)(1/2-1/\theta)}.$$

Theorem 6.24 has been proved in [396] in a more general and non-periodic situation. The author used discretization techniques (wavelet isomorphisms), entropy results for finite dimensional spaces, and complex interpolation to obtain the above result.

If either $r > 1/p$ and $\theta \geq p$ or $r > (1/p - 1/q)_+$ and $\theta \geq \min\{q, 2\}$ the result in Theorem 6.24 has been shown by Dinh Dũng [96]. One should keep in mind that under these parameter assumptions it is possible to obtain that $r + 1/2 - 1/\theta < 0$. Of course, the exponent of the logarithm can not be negative. Inspecting the proof of [96, Thm. 8] it turns out that Dinh Dũng proved a slightly weaker upper bound in the situation $r > (1/p - 1/q)_+$ and $\theta \geq \min\{q, 2\}$, namely

$$\epsilon_n(\mathbf{B}_{p,\theta}^r, L_q) \lesssim n^{-r} (\log n)^{(d-1)(r+1/\min\{q,2\}-1/\theta)}.$$

Very recently, in the paper [374] the approximation classes $\mathbf{W}_p^{a,b}$ have been studied from the point of view of the entropy numbers. We do not give a detailed definition of these classes here, we only note that these classes are defined in a way similar to the classes $\mathbf{W}_A^{a,b}$ (see Lemma 7.16 below) with $\|\cdot\|_A$ replaced by $\|\cdot\|_p$. The paper [374] develops a new method of proving the upper bounds for the entropy numbers. This method, which is based on the general Theorem 6.9, allows us to prove all known upper bounds for classes \mathbf{W}_p^r .

6.4 Entropy numbers and the Small Ball Problem

We already pointed out that the case $r = 1$, $p = 2$, $q = \infty$ is equivalent to the Small Ball Problem from probability theory.

We discuss related results in detail. Consider the centered Gaussian process $\mathbf{B}_d := (B_{\mathbf{x}})_{\mathbf{x} \in [0,1]^d}$ with covariance

$$\mathbb{E}(B_{\mathbf{x}} B_{\mathbf{y}}) = \prod_{i=1}^d \min\{x_i, y_i\}, \quad \mathbf{x} = (x_1, \dots, x_d), \quad \mathbf{y} = (y_1, \dots, y_d).$$

This process is called Brownian sheet. It is known that the sample paths of \mathbf{B}_d are almost surely continuous. We consider them as random elements of the space $C([0, 1]^d)$. The small ball problem is the problem of the asymptotic behavior of the small ball probabilities

$$\mathbb{P}\left(\sup_{\mathbf{x} \in [0, 1]^d} |B_{\mathbf{x}}| \leq \epsilon\right)$$

as ϵ tends to zero. We introduce notation

$$\phi(\epsilon) := -\ln \mathbb{P}\left(\sup_{\mathbf{x} \in [0, 1]^d} |B_{\mathbf{x}}| \leq \epsilon\right).$$

The following relation is a fundamental result of probability theory: for $d = 2$ and $\epsilon < 1/2$

$$\phi(\epsilon) \asymp \epsilon^{-2} (\ln(1/\epsilon))^3. \quad (6.11)$$

The upper bound in (6.11) was obtained by Lifshits and Tsirelson [219] and by Bass [21]. The lower bound in (6.11) was obtained by Talagrand [313].

Kuelbs and Li [206] discovered the fact that there is a tight relationship between $\phi(\epsilon)$ and the entropy $H_\epsilon(\mathbf{W}_2^1, L_\infty)$. We note that they considered the general setting of a Gaussian measure on a Banach space. We only formulate a particular result of our interest in terms convenient for us. Roughly speaking (i.e. under the additional condition $\phi(\epsilon) \sim \phi(\epsilon/2)$), they proved the equivalence relations: for any d

$$\begin{aligned} \phi(\epsilon) \lesssim \epsilon^{-2} (\ln(1/\epsilon))^\beta &\iff \epsilon_n(\mathbf{W}_2^1, L_\infty) \lesssim n^{-1} (\ln n)^{\beta/2}; \\ \phi(\epsilon) \gtrsim \epsilon^{-2} (\ln(1/\epsilon))^\beta &\iff \epsilon_n(\mathbf{W}_2^1, L_\infty) \gtrsim n^{-1} (\ln n)^{\beta/2}. \end{aligned}$$

These relations and (6.11) imply for $d = 2$

$$\epsilon_n(\mathbf{W}_2^1(\mathbb{T}^2), L_\infty(\mathbb{T}^2)) \asymp n^{-1} (\ln n)^{3/2}. \quad (6.12)$$

The proof of the most difficult part of (6.11) – the lower bound – is based on a special inequality for the Haar polynomials proved by Talagrand [313] (see [355] for a simple proof). The reader can find this inequality in Subsection 2.6 of Section 2. We note that the lower bound in (6.12) can be deduced directly from the corresponding Small Ball Inequality (2.26). However, this way does not work for deducing the lower bound in (6.5) for general r . This difficulty was overcome in [354] by proving the Small Ball Inequality for the trigonometric system (see Subsection 2.6 of Section 2, inequality (2.27)).

The other way around, results on entropy numbers can be used to estimate the small ball probabilities. From the relation in (6.6) one obtains the bounds

$$\epsilon^{-2} (\ln(1/\epsilon))^{2d-2} \lesssim \phi(\epsilon) \lesssim \epsilon^{-2} (\ln(1/\epsilon))^{2d-1}, \quad (6.13)$$

see also [218]. Note, that there is so far no “pure” probabilistic proof for the upper bound. By the recent improvements on the Small Ball Inequality (see the remark after (2.29) in Subsection 2.6 above) one can slightly improve the lower bound in (6.13) to get

$$\epsilon^{-2} (\ln(1/\epsilon))^{2d-2+\delta} \lesssim \phi(\epsilon) \lesssim \epsilon^{-2} (\ln(1/\epsilon))^{2d-1},$$

for some $0 < \delta = \delta(d) < 1$.

6.5 Open problems

We presented historical comments and a discussion in the above text of Section 6. We formulate here the most important open problems on the entropy numbers of classes of functions with mixed smoothness.

Open problem 6.1. Find the order of $\epsilon_n(\mathbf{W}_\infty^r, L_\infty)$ in the case $d = 2$ and $0 < r \leq 1/2$.

Open problem 6.2. For $d > 2$ find the order of $\epsilon_n(\mathbf{W}_1^r, L_q)$, $1 \leq q \leq \infty$.

Open problem 6.3. For $d > 2$ find the order of $\epsilon_n(\mathbf{W}_p^r, L_\infty)$ and $\epsilon_n(\mathbf{H}_p^r, L_\infty)$, $1 \leq p \leq \infty$, $r > 1/p$.

Open problem 6.4. Find the correct order of the entropy numbers $\epsilon_n(\mathbf{B}_{p,\theta}^r, L_2)$ in case of small smoothness $0 < \theta < p \leq 2$ and $1/p - 1/2 < r \leq 1/\theta - 1/2$.

7 Best m -term approximation

7.1 Introduction

The last two decades have seen great successes in studying nonlinear approximation which was motivated by numerous applications. Nonlinear approximation is important in applications because of its concise representations and increased computational efficiency. Two types of nonlinear approximation are frequently employed in applications. Adaptive methods are used in PDE solvers, while m -term approximation, considered here, is used in image/signal/data processing, as well as in the design of neural networks. Another name for m -term approximation is *sparse approximation*.

The fundamental question of nonlinear approximation is how to devise good constructive methods (algorithms) of nonlinear approximation. This problem has two levels of nonlinearity. The first level of nonlinearity is m -term approximation with regard to bases. In this problem one can use the unique function expansion with regard to a given basis to build an approximant. Nonlinearity enters by looking for m -term approximants with terms (i.e. basis elements in approximant) allowed to depend on a given function. Since the elements of the basis used in the m -term approximation are allowed to depend on the function being approximated, this type of approximation is very efficient. On the second level of nonlinearity, we replace a basis by a more general system which is not necessarily minimal (for example, redundant system, dictionary). This setting is much more complicated than the first one (bases case), however, there is a solid justification of importance of redundant systems in both theoretical questions and in practical applications. We only give here a brief introduction to this important area of research and refer the reader to the book [368] for further results.

One of the major questions in approximation (theoretical and numerical) is: what is an optimal method? We discuss here this question in a theoretical setting with the only criterion of quality of approximating method its accuracy. One more important point in the setting of optimization problem is to specify a set of methods over which we are going to optimize. Most of the problems which approximation theory deals with are of this nature. Let us give some examples from classical approximation theory. These examples will help us to motivate the question we are studying in this section.

Example 1. When we are searching for n -th best trigonometric approximation of a given function we are optimizing in the sense of accuracy over the subspace of trigonometric polynomials of degree n .

Example 2. When we are solving the problem on Kolmogorov's n -width for a given function class we are optimizing in the sense of accuracy for a given class over all subspaces of dimension n .

Example 3. When we are finding best m -term approximation of a given function with regard to a given system of functions (dictionary) we are optimizing over all m -dimensional subspaces spanned by elements from a given dictionary.

Example 2 is a development of Example 1 in the sense that in Example 2 we are looking for an optimal n -dimensional subspace instead of being confined to a given one (trigonometric polynomials of degree n). Example 3 is a nonlinear analog of Example 1, where instead of trigonometric system we take a dictionary \mathcal{D} and allow approximating elements from \mathcal{D} to depend on a function. In [360] we made some steps in a direction of developing Example 3 to a setting which is a nonlinear analog of Example 2. In other words, we want to optimize over some sets of dictionaries. We discuss two classical structural properties of dictionaries:

1. Orthogonality;
2. Tensor product structure (multivariate case).

Denote by \mathcal{D} a dictionary in a Banach space X and by

$$\sigma_m(f, \mathcal{D})_X := \inf_{\substack{g_i \in \mathcal{D}, c_i \\ i=1, \dots, m}} \|f - \sum_{i=1}^m c_i g_i\|_X$$

best m -term approximation of f with regard to \mathcal{D} . For a function class $F \subset X$ and a collection \mathbb{D} of dictionaries we consider

$$\sigma_m(F, \mathcal{D})_X := \sup_{f \in F} \sigma_m(f, \mathcal{D})_X \quad ,$$

$$\sigma_m(F, \mathbb{D})_X := \inf_{\mathcal{D} \in \mathbb{D}} \sigma_m(F, \mathcal{D})_X \quad .$$

Thus the quantity $\sigma_m(F, \mathbb{D})_X$ gives the sharp lower bound for best m -term approximation of a given function class F with regard to any dictionary $\mathcal{D} \in \mathbb{D}$.

Denote by \mathbb{O} the set of all orthonormal dictionaries defined on a given domain. Kashin [194] proved that for the class $H^{r, \alpha}$, $r = 0, 1, \dots$, $\alpha \in [0, 1]$, of univariate functions such that

$$\|f\|_\infty + \|f^{(r)}\|_\infty \leq 1 \quad \text{and} \quad |f^{(r)}(x) - f^{(r)}(y)| \leq |x - y|^\alpha, \quad x, y \in [0, 1]$$

we have

$$\sigma_m(H^{r, \alpha}, \mathbb{O})_{L_2} \geq C(r, \alpha) m^{-r-\alpha}. \quad (7.1)$$

It is well known that in the case $\alpha \in (0, 1)$ the class $H^{r, \alpha}$ is equivalent to the class $H_\infty^{r+\alpha}$. In the case $\alpha = 1$ the class $H^{r, 1}$ is close to the class W_∞^{r+1} . It is interesting to remark that we cannot prove anything like (7.1) with L_2 replaced by L_p , $p < 2$. We proved (see [195]) that there exists $\Phi \in \mathbb{O}$ such that for any $f \in L_1(0, 1)$ we have $\sigma_1(f, \Phi)_{L_1} = 0$. It is pointed out in [360] that the proof from [195] also works for L_p , $p < 2$, instead of L_1 :

Remark 7.1. *For any $1 \leq p < 2$ there exists a complete in $L_2(0, 1)$ orthonormal system Φ such that for each $f \in L_p(0, 1)$ we have $\sigma_1(f, \Phi)_{L_p} = 0$.*

This remark means that to obtain nontrivial lower bounds for $\sigma_m(f, \Phi)_{L_p}$, $p < 2$, we need to impose additional restrictions on $\Phi \in \mathbb{O}$. Some ways of imposing restrictions were discussed in [195] and [360].

7.2 Orthogonal bases

We discuss approximation of multivariate functions. It is convenient for us to present results in the periodic case. We consider classes of functions with bounded mixed derivative \mathbf{W}_p^r and classes with restriction of Lipschitz type on mixed difference \mathbf{H}_p^r and $\mathbf{B}_{p, \theta}^r$. These classes are well known (see for instance [352]) for their importance in numerical integration, in finding universal methods for approximation of functions of several variables, in the average case setting of approximation problems for the spaces equipped with the Wiener sheet measure (see [405]) and in other problems. It is proved in [360] that

$$\sigma_m(\mathbf{H}_p^r, \mathbb{O})_2 \gtrsim m^{-r} (\log m)^{(d-1)(r+1/2)}, \quad 1 \leq p < \infty, \quad (7.2)$$

$$\sigma_m(\mathbf{W}_p^r, \mathbb{O})_2 \gtrsim m^{-r} (\log m)^{(d-1)r}, \quad 1 \leq p < \infty. \quad (7.3)$$

It is also proved in [360] that the orthogonal basis \mathcal{U}^d which we construct below provides optimal upper estimates (like (7.2) and (7.3)) in best m -term approximation of the classes \mathbf{H}_p^r and \mathbf{W}_p^r in the L_q -norm, $2 \leq q < \infty$. Moreover, we proved there that for all $1 < p, q < \infty$ the

order of best m -term approximation $\sigma_m(\mathbf{H}_p^r, \mathcal{U}^d)_{L_q}$ and $\sigma_m(\mathbf{W}_p^r, \mathcal{U}^d)_{L_q}$ can be achieved by a greedy type algorithm $G^q(\cdot, \mathcal{U}^d)$. Assume a given system Ψ of functions ψ_I indexed by dyadic intervals can be enumerated in such a way that $\{\psi_{I_j}\}_{j=1}^\infty$ is a basis for L_q . Then we define the Thresholding Greedy Algorithm (TGA) $G^q(\cdot, \Psi)$ as follows. Let

$$f = \sum_{j=1}^{\infty} c_{I_j}(f, \Psi) \psi_{I_j}$$

and

$$c_I(f, q, \Psi) := \|c_I(f, \Psi) \psi_I\|_q.$$

Then $c_I(f, q, \Psi) \rightarrow 0$ as $|I| \rightarrow 0$. Denote Λ_m a set of m dyadic intervals I such that

$$\min_{I \in \Lambda_m} c_I(f, q, \Psi) \geq \max_{J \notin \Lambda_m} c_J(f, q, \Psi).$$

We define $G^q(\cdot, \Psi)$ by formula

$$G_m^q(f, \Psi) := \sum_{I \in \Lambda_m} c_I(f, \Psi) \psi_I.$$

The question of constructing a procedure (theoretical algorithm) which realizes (in the sense of order) the best possible accuracy is a very important one and we discuss it in detail in this section. Let $A_m(\cdot, \mathcal{D})$ be a mapping which maps each $f \in X$ to a linear combination of m elements from a given dictionary \mathcal{D} . Then the best we can hope for with this mapping is to have for each $f \in X$

$$\|f - A_m(f, \mathcal{D})\|_X = \sigma_m(f, \mathcal{D})_X \quad (7.4)$$

or a little weaker

$$\|f - A_m(f, \mathcal{D})\|_X \leq C(\mathcal{D}, X) \sigma_m(f, \mathcal{D})_X. \quad (7.5)$$

There are some known trivial and nontrivial examples when (7.4) holds in a Hilbert space X . We do not touch this kind of relations here. Concerning (7.5) it is proved in [361] that for any basis Ψ which is L_p -equivalent to the univariate Haar basis we have

$$\|f - G_m^q(f, \Psi)\|_q \leq C(p) \sigma_m(f, \Psi)_q, \quad 1 < q < \infty. \quad (7.6)$$

However, as it is shown in [359] and [360], the inequality (7.6) does not hold for particular dictionaries with tensor product structure.

We define the system $\mathcal{U} := \{U_I\}$ in the univariate case. Denote

$$\begin{aligned} U_n^+(x) &:= \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^n x} - 1}{e^{ix} - 1}, \quad n = 0, 1, 2, \dots; \\ U_{n,k}^+(x) &:= e^{i2^n x} U_n^+(x - 2\pi k 2^{-n}), \quad k = 0, 1, \dots, 2^n - 1; \\ U_{n,k}^-(x) &:= e^{-i2^n x} U_n^+(-x + 2\pi k 2^{-n}), \quad k = 0, 1, \dots, 2^n - 1. \end{aligned}$$

It will be more convenient for us to normalize in L_2 the system of functions $\{U_{m,k}^+, U_{n,k}^-\}$ and enumerate it by dyadic intervals. We write

$$\begin{aligned} U_I(x) &:= 2^{-n/2} U_{n,k}^+(x) \quad \text{with} \quad I = [(k+1/2)2^{-n}, (k+1)2^{-n}); \\ U_I(x) &:= 2^{-n/2} U_{n,k}^-(x) \quad \text{with} \quad I = [k2^{-n}, (k+1/2)2^{-n}); \end{aligned}$$

and

$$U_{[0,1)}(x) := 1.$$

Denote

$$\begin{aligned} D_n^+ &:= \{I : I = [(k + 1/2)2^{-n}, (k + 1)2^{-n}), \quad k = 0, 1, \dots, 2^n - 1\}; \\ D_n^- &:= \{I : I = [k2^{-n}, (k + 1/2)2^{-n}), \quad k = 0, 1, \dots, 2^n - 1\}; \end{aligned}$$

and

$$D := \cup_{n \geq 0} (D_n^+ \cup D_n^-) \cup D_0$$

with $D_0 := [0, 1)$. In the multivariate case of $\mathbf{x} = (x_1, \dots, x_d)$ we define the system \mathcal{U}^d as the tensor product of the univariate systems \mathcal{U} . Let $I = I_1 \times \dots \times I_d$, $I_j \in D$, $j = 1, \dots, d$, then

$$U_I(\mathbf{x}) := \prod_{j=1}^d U_{I_j}(x_j).$$

We have for instance (see [360])

$$\sup_{f \in L_q} \|f - G_m^q(f, \mathcal{U}^d)\|_{L_q} / \sigma_m(f, \mathcal{U}^d)_{L_q} \gtrsim (\log m)^{(d-1)|1/2-1/q|}. \quad (7.7)$$

The inequality (7.7) shows that using the algorithm $G^q(\cdot, \mathcal{U}^d)$ we lose for sure for some functions $f \in L_q$, $q \neq 2$. In light of (7.7) the results of [360] look encouraging for using $G^q(\cdot, \mathcal{U}^d)$.

Theorem 7.2. *Define*

$$r(W, p, q) = \begin{cases} \max\{1/p, 1/2\} - 1/q & : \quad q \geq 2, \\ (\max\{2/q, 2/p\} - 1)/q & : \quad p < 2. \end{cases}.$$

Then for $1 < p, q < \infty$ and $r > r(W, p, q)$ we have

$$\sup_{f \in \mathbf{W}_p^r} \|f - G_m^q(f, \mathcal{U}^d)\|_q \asymp \sigma_m(\mathbf{W}_p^r, \mathcal{U}^d)_q \asymp m^{-r} (\log m)^{(d-1)r}. \quad (7.8)$$

Theorem 7.3. *Define*

$$r(H, p, q) = \begin{cases} (1/p - 1/q)_+ & : \quad q \geq 2, \\ (\max\{2/q, 2/p\} - 1)/q & : \quad q < 2. \end{cases}.$$

Then for $1 < p, q < \infty$ and $r > r(H, p, q)$ we have

$$\sup_{f \in \mathbf{H}_p^r} \|f - G_m^q(f, \mathcal{U}^d)\|_q \asymp \sigma_m(\mathbf{H}_p^r, \mathcal{U}^d)_q \asymp m^{-r} (\log m)^{(d-1)(r+1/2)}. \quad (7.9)$$

Comparing (7.9) with (7.2) and (7.8) with (7.3), we conclude that the dictionary \mathcal{U}^d is the best (in the sense of order) among all orthogonal dictionaries for m -term approximation of the classes \mathbf{H}_p^r and \mathbf{W}_p^r in L_q where $1 < p < \infty$ and $2 \leq q < \infty$. The dictionary \mathcal{U}^d has one more important feature. The near best m -term approximation of functions from \mathbf{H}_p^r and \mathbf{W}_p^r in the L_q -norm can be realized by the simple greedy type algorithm $G^q(\cdot, \mathcal{U}^d)$ for all $1 < p, q < \infty$. For further results in this direction for the system \mathcal{U}^d we refer to [13], [14].

It is known that the system \mathcal{U}^d and its analog built on the base of the de la Vallée Poussin kernels instead of the Dirichlet kernels play important role in the bilinear approximation [339]

(for bilinear approximation see subsection 7.4 below). The de la Vallée Poussin kernels are especially important when we deal with either L_1 or L_∞ spaces. This setting has been considered by Dinh Dũng in [95, 97]. Instead of the system \mathcal{U}^d he considered translates of the de la Vallée Poussin kernels \mathcal{V}^d . In contrast to \mathcal{U}^d this one is linearly dependent (redundant). However, it admits similar discretization techniques as orthonormal wavelet bases. In addition, the spaces studied slightly differ from the ones considered by Temlyakov.

Theorem 7.4. *Let $1 < p, q < \infty$, $0 < \theta \leq \infty$ and $r > 0$.*

(i) *It holds*

$$\sigma_m(\mathbf{B}_{p,\theta}^r, \mathcal{V}^d)_q \gtrsim m^{-r} (\log m)^{(d-1)(r+1/2-1/\theta)}.$$

(ii) *If $r > (1/p - 1/q)_+$ and $\theta \geq \min\{q, 2\}$. Then*

$$\sigma_m(\mathbf{B}_{p,\theta}^r, \mathcal{V}^d)_q \lesssim m^{-r} (\log m)^{(d-1)(r+1/\min\{q,2\}-1/\theta)}.$$

(iii) *Let $r > \max\{0, 1/p - 1/q, 1/p - 1/2\}$ and $2 \leq \theta \leq \infty$. Then it holds*

$$\sigma_m(\mathbf{B}_{p,\theta}^r, \mathcal{V}^d)_q \asymp m^{-r} (\log m)^{(d-1)(r+1/2-1/\theta)}.$$

(iv) *Let $r > 1/p$. Then*

$$\sigma_m(\mathbf{B}_{p,\theta}^r, \mathcal{V}^d)_q \lesssim m^{-r} (\log m)^{(d-1)(r+1/2-1/\max\{p,\theta\})}.$$

Dinh Dũng essentially studied embeddings between Besov spaces and L_q . As a consequence of well-known embeddings he obtained also the following result for the \mathbf{W} spaces which has to be compared with Theorem 7.2 above.

Theorem 7.5. *Let $1 < p, q < \infty$ and $r > \max\{0, 1/p - 1/q, 1/p - 1/2, 1/2 - 1/q\}$. Then*

$$\sigma_m(\mathbf{W}_p^r, \mathcal{V}^d)_q \asymp m^{-r} (\log m)^{(d-1)r}.$$

Let now $\Phi = \{\psi_I\}_I$ be a tensorized orthonormal wavelet basis (indexed by a dyadic parallelepiped I) with sufficient smoothness, decay and vanishing moments. In the non-periodic setting the estimates for σ_m in Theorems 7.2, 7.3, 7.4 can be also observed in a non-periodic setting, see [360, 164, 165, 166] using the wavelet basis Φ as a dictionary. Here we get the sharp bounds.

Theorem 7.6. *Let $1 < p, q < \infty$ and $0 < \theta \leq \infty$*

(i) *Let $r > \max\{1/p - 1/q, 1/\min\{p, \theta\} - 1/\max\{q, 2\}, 0\}$ then*

$$\sigma_m(\mathbf{B}_{p,\theta}^r, \Phi)_q \asymp m^{-r} (\log m)^{(d-1)(r+1/2-1/\theta)_+}.$$

(ii) *Let $r > (1/p - 1/q)_+$ then*

$$\sigma_m(\mathbf{W}_p^r, \Phi)_q \asymp m^{-r} (\log m)^{(d-1)r}.$$

Note, that contrary to Theorems 7.2 and 7.3 Theorem 7.6 does not address the greedy approximation. As a result, the assumptions in Theorem 7.6 are (so far) the weakest compared to Theorems 7.2, 7.4, 7.5.

Let us also address a small smoothness effect in this framework (observed by Hansen, Sickel [166, Cor. 5.11]) according to the parameter situation already mentioned after Theorem 6.24 in the Entropy section (see also Open problem 6.4). In fact, in case $0 < \theta < p \leq 2$ and $1/p - 1/2 < r \leq 1/\theta - 1/2$ we have

$$\sigma_m(\mathbf{B}_{p,\theta}^r, \Phi)_2 \asymp m^{-r}. \quad (7.10)$$

The bound is a simple consequence of recent results [164, 166] on best m -term approximation for non-compact embeddings $\mathbf{W}_p^r \hookrightarrow L_q$ if $1 < p < q < \infty$ and $r = 1/p - 1/q$, and $\mathbf{B}_{p,p}^r \hookrightarrow L_q$ if $0 < p \leq \max\{p, 1\} < q < \infty$ and $r = 1/p - 1/q$, see Theorems 9.9, 9.11 below. Clearly, in the above parameter situation we may use the trivial embedding $\mathbf{B}_{p,\theta}^r \hookrightarrow \mathbf{B}_{p^*,p^*}^r$, where p^* is chosen such that $r = 1/p^* - 1/2$ and hence we may use the non-compact embedding result (9.4) below. Note, that in (7.10) the typical logarithmic factor with the d power is not present. From the viewpoint of high-dimensional approximation this is an important phenomenon. Let us also refer to the discussions in Section 10 and Subsection 7.5 below (after Theorem 7.12).

In the recent paper [49] best m -term approximation with respect to the tensor product Faber-Schauder system \mathcal{F}^d , see Subsection 5.5, has been addressed. Via a greedy type algorithm one can prove for $1 < p < q \leq \infty$, $\max\{1/p, 1/2\} < r < 2$ that

$$\sigma_m(\mathbf{W}_p^r, \mathcal{F}^d)_q \lesssim m^{-r} (\log m)^{(d-1)(r+1/2)}.$$

Note, that every coefficient in the Faber-Schauder expansion is given by a linear combination of (discrete) function values (5.38). Hence, the best m -term approximant is build from $c(d)m$ properly chosen function values and corresponding hat functions.

7.3 Some related problems

Approximations with respect to the basis \mathcal{U}^d and with respect to the trigonometric system are useful in the following two fundamental problems.

Bilinear approximation. For a function $f(\mathbf{x}, \mathbf{y}) \in L_{p_1, p_2}$, $\mathbf{x} = (x_1, \dots, x_a)$, $\mathbf{y} = (y_1, \dots, y_a)$ we define the best bilinear approximation as follows

$$\tau_m(f)_{p_1, p_2} := \inf_{\substack{u_i(\mathbf{x}), v_i(\mathbf{y}) \\ i=1, \dots, m}} \left\| f(\mathbf{x}, \mathbf{y}) - \sum_{i=1}^m u_i(\mathbf{x}) v_i(\mathbf{y}) \right\|_{p_1, p_2}.$$

where $\|\cdot\|_{p_1, p_2}$ denotes the mixed norm: L_{p_1} in \mathbf{x} on \mathbb{T}^a and in L_{p_2} in \mathbf{y} on \mathbb{T}^a .

The classical result, obtained by E. Schmidt [288], gives the following theorem.

Theorem 7.7. *Let $K(\mathbf{x}, \mathbf{y}) \in L_{2,2}$. Then*

$$\tau_m(K)_{2,2} = \left(\sum_{j=m+1}^{\infty} \lambda_j(K) \right)^{1/2},$$

where $\{\lambda_j(K)\}_{j=1}^{\infty}$ is a non-increasing sequence of eigenvalues of the operator K^*K with K being an integral operator with the kernel $K(\mathbf{x}, \mathbf{y})$.

It is clear that for a periodic function $f \in L_q(\mathbb{T}^a)$, one has

$$\tau_m(f(\mathbf{x} - \mathbf{y}))_{q,\infty} \leq \sigma_m(f, \mathcal{T}^a)_q.$$

This observation and some known results on the m -term approximations allowed us to prove sharp upper bounds for the best bilinear approximation (see [336] and [340]). We note here that it easily follows from the definition of classes \mathbf{W}_p^r that

$$\lambda_m(\mathbf{W}_1^r, L_q) \leq \tau_m(F_r(\mathbf{x} - \mathbf{y}))_{q,\infty} \leq \sigma_m(F_r, \mathcal{T})_q.$$

In the case $d = 2$ ($a = 1$) it is the classical problem of bilinear approximation. It turned out that the best m -term approximation with respect to the trigonometric system gives the best bilinear

approximation in the sense of order for some classes of functions. In the case of approximation in the L_2 -space the bilinear approximation problem is closely related to the problem of singular value decomposition (also called Schmidt expansion) of the corresponding integral operator with the kernel $f(x_1, x_2)$ (see Theorem 7.7 above). There are known results on the rate of decay of errors of best bilinear approximation in L_{p_1, p_2} under different smoothness assumptions on f . We only mention some known results for classes of functions which are studied in this paper. The problem of estimating $\tau_m(f)_2$ in case $d = 2$ (best m -term bilinear approximation $\tau_m(f)$ in L_2) is a classical one and was considered for the first time by E. Schmidt [288] in 1907. For many function classes \mathbf{F} an asymptotic behavior of $\tau_m(\mathbf{F})_p := \sup_{f \in \mathbf{F}} \tau_m(f)_{p,p}$ is known. For instance, the relation

$$\tau_m(\mathbf{W}_p^r)_q \asymp m^{-2r+(1/p-\max\{1/2, 1/q\})_+}$$

for $r > 1$ and $1 \leq p \leq q \leq \infty$ follows from more general results in [341].

Tensor product approximation. For a function $f(x_1, \dots, x_d)$ denote

$$\Theta_m(f)_X := \inf_{\substack{\{u_j^i\} \\ j=1, \dots, m \\ i=1, \dots, d}} \left\| f(x_1, \dots, x_d) - \sum_{j=1}^m \prod_{i=1}^d u_j^i(x_i) \right\|_X.$$

In the case $d > 2$ almost nothing is known for the tensor product approximation. There is (see [343]) an upper estimate in the case $q = p = 2$

$$\Theta_m(\mathbf{W}_2^r)_2 := \sup_{f \in \mathbf{W}_2^r} \Theta_m(f)_{L_2} \lesssim m^{-rd/(d-1)}.$$

For recent results in this direction see [26] and [291].

7.4 A comment on Stechkin's lemma

The following simple lemma (see, for instance, [340, p. 92]) turns out to be very useful in nonlinear approximation. Note, that here the case $p < 1$ is particularly important and leads to better convergence rates.

Lemma 7.8. *Let $a_1 \geq a_2 \geq \dots \geq a_M \geq 0$ and $0 < p \leq q \leq \infty$. Then for all $m < M$ one has*

$$\left(\sum_{k=m}^M a_k^q \right)^{1/q} \leq m^{-\beta} \left(\sum_{k=1}^M a_k^p \right)^{1/p}, \quad \beta := 1/p - 1/q.$$

Proof. Denote

$$A := \left(\sum_{k=1}^M a_k^p \right)^{1/p}.$$

Monotonicity of $\{a_k\}$ implies

$$ma_m^p \leq A^p \quad \text{and} \quad a_m \leq Am^{-1/p}.$$

Then

$$\sum_{k=m}^M a_k^q \leq a_m^{q-p} \sum_{k=m}^M a_k^p \leq a_m^{q-p} A^p.$$

The above two inequalities imply Lemma 7.8. ■

For the first time a version of Lemma 7.8 for $1 \leq p \leq 2, q = 2$, was proved and used in nonlinear approximation in [336]. In the case $1 \leq p \leq q \leq \infty$ Lemma 7.8 is proved in [340]. The same proof gives Lemma 7.8 (see above) for $0 < p \leq q \leq \infty$.

It has been observed recently that the constant in Lemma 7.8 is actually better than stated ($c_{p,q} = 1$ in Lemma 7.8). Using tools from convex optimization one can prove the following stronger version of Lemma 7.8, see [123], [125, Thm. 2.5].

Lemma 7.9. *Let $a_1 \geq a_2 \geq \dots \geq a_M \geq 0$ and $0 < p < q \leq \infty$. Then for all $m < M$ one has*

$$\left(\sum_{k=m+1}^M a_k^q \right)^{1/q} \leq c_{p,q} m^{-\beta} \left(\sum_{k=1}^M a_k^p \right)^{1/p}, \quad \beta := 1/p - 1/q,$$

where

$$c_{p,q} := \left[\left(\frac{p}{q} \right)^{p/q} \left(1 - \frac{p}{q} \right)^{1-p/q} \right]^{1/p} \leq 1.$$

Note, that in case $q = 2$ and $p = 1$ we obtain $c_{1,2} = 1/2$. This result is useful if the p is not determined in advance. In fact, in various situations it is possible to minimize the right-hand side in Lemma 7.9 by choosing p depending on m via balancing the quantities involved in the right-hand side (including $c_{p,q}$).

Lemma 7.8 and its versions are used in many papers dealing with nonlinear approximation and, more recently, the approximation of high-dimensional parametric elliptic partial differential equations as in (10.6) below, see e.g. [60, Lem. 3.6], [62, (3.13)], [9, (21)], [381] and Subsection 10.4. This is related to the novel field of *uncertainty quantification*. In some of these papers it is called Stechkin's lemma. We make a historical remark in this regard. As far as we know S.B. Stechkin did not formulate Lemma 7.8 even in a special case. Studying absolute convergence of Fourier series, Stechkin [310] proved the following lemma.

Lemma 7.10. (The Stechkin lemma) *Let $a_1 \geq a_2 \geq \dots$ be a sequence of nonnegative numbers. Then the following inequalities hold*

$$\frac{1}{2} \sum_{n=1}^{\infty} n^{-1/2} \left(\sum_{k=n}^{\infty} a_k^2 \right)^{1/2} \leq \sum_{k=1}^{\infty} a_k \leq \frac{2}{\sqrt{3}} \sum_{n=1}^{\infty} n^{-1/2} \left(\sum_{k=n}^{\infty} a_k^2 \right)^{1/2}.$$

Classically (see [310], [76], [69, Thm. 4], [258, Ex. 1 in 3.2]) authors considered the following generalization of Lemma 7.10, which is stronger in a certain sense than Lemma 7.8. These generalizations are used to characterize nonlinear approximation spaces, see for instance Theorem 9.11 below. For any $x \in \ell_q$ we denote

$$\sigma_m(x)_q := \inf \{ \|x - z\|_q : z \in \ell_q, m\text{-sparse} \}$$

Lemma 7.11. *Let $0 < p < q \leq \infty$. Then $x \in \ell_q$ belongs to ℓ_p if and only if*

$$\left[\sum_{m=1}^{\infty} \left(m^{1/p-1/q} \sigma_m(x)_q \right)^p \frac{1}{m} \right]^{1/p} < \infty. \quad (7.11)$$

Then (7.11) is an equivalent quasi-norm in ℓ_p .

Lemma 7.11 means in particular that for any $x \in \ell_p$ one has

$$\left[\sum_{m=1}^{\infty} \left[m^{1/p-1/q} \sigma_m(x)_q \right]^p \frac{1}{m} \right]^{1/p} \leq C(p, q) \|x\|_p. \quad (7.12)$$

One can derive Lemma 7.8 (with the constant 1 replaced by $C(p, q)$) from (7.12) by an argument, which is very similar to the direct proof of Lemma 7.8.

7.5 Sparse trigonometric approximation

Sparse trigonometric approximation of periodic functions began by the paper of S.B. Stechkin [310], who used it in the criterion for absolute convergence of trigonometric series. R.S. Ismagilov [180] found nontrivial estimates for m -term approximation of functions with singularities of the type $|x|$ and gave interesting and important applications to the widths of Sobolev classes. He used a deterministic method based on number theoretical constructions. His method was developed by V.E. Maiorov [227], who used a method based on Gaussian sums. Further strong results were obtained in [75] with the help of a non-constructive result from finite dimensional Banach spaces due to E.D. Gluskin [145]. Other powerful non-constructive method, which is based on a probabilistic argument, was used by Y. Makovoz [228] and by E.S. Belinskii [33] (see the book [385] for a detailed description of the technique). Different methods were created in [336], [195], [358], [369] for proving lower bounds for function classes. It was discovered in [81] and [366] that greedy algorithms can be used for constructive m -term approximation with respect to the trigonometric system. We demonstrated in [371] how greedy algorithms can be used to prove optimal or best known upper bounds for m -term approximation of classes of functions with mixed smoothness. It is a simple and powerful method of proving upper bounds. The reader can find a detailed study of m -term approximation of classes of functions with mixed smoothness, including small smoothness, in the paper [271] by A.S. Romanyuk and in recent papers [371], [372]. We note that in the case $2 < q < \infty$ the upper bounds in [271] are not constructive.

The following two theorems are from [371]. We use the notation $\beta := \beta(p, q) := 1/p - 1/q$ and $\eta := \eta(p) := 1/p - 1/2$. In the case of trigonometric system \mathcal{T}^d we drop it from the notation:

$$\sigma_m(\mathbf{F})_q := \sigma_m(\mathbf{F}, \mathcal{T}^d)_q.$$

Theorem 7.12. *We have*

$$\sigma_m(\mathbf{W}_p^r)_q \asymp \begin{cases} m^{-r+\beta}(\log m)^{(d-1)(r-2\beta)}, & 1 < p \leq q \leq 2, \quad r > 2\beta, \\ m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)}, & 1 < p \leq 2 \leq q < \infty, \quad r > 1/p, \\ m^{-r}(\log m)^{r(d-1)}, & 2 \leq p \leq q < \infty, \quad r > 1/2. \end{cases}$$

The third line of Theorem 7.12 combined with (7.3) shows that the trigonometric system is an optimal in the sense of order orthonormal system for the m -term approximation of the classes \mathbf{W}_p^r in L_q for $2 \leq p, q < \infty$. The case $1 < p \leq q \leq 2$ in Theorem 7.12, which corresponds to the first line, was proved in [336] (see also [340], Ch.4). The proofs from [336] and [340] are constructive.

Comparing Theorems 7.12 and 7.26 (below) with Theorem 4.4 we conclude that in the case $1 < q \leq p < \infty$ we have

$$\sigma_m(\mathbf{W}_p^r)_q \asymp E_{Q_n}(\mathbf{W}_p^r)_q, \quad m \asymp |Q_n| \asymp 2^n n^{d-1}.$$

In the case $1 < p < q < \infty$ we have

$$\sigma_m(\mathbf{W}_p^r)_q = o(E_{Q_n}(\mathbf{W}_p^r)_q), \quad m \asymp |Q_n| \asymp 2^n n^{d-1}.$$

Comparing Theorems 7.12 and 7.26 (below) with Theorem 7.2 we see that the trigonometric system \mathcal{T}^d is as good as the wavelet type system \mathcal{U}^d for m -term approximation in the range of parameters $1 < q \leq p < \infty$ and $2 \leq p \leq q < \infty$ and for other parameters from $1 < p, q < \infty$ the wavelet type system \mathcal{U}^d provides better m -term approximation than the trigonometric system \mathcal{T}^d . Similar phenomenon holds for the \mathbf{H} classes as well. Probably, this phenomenon is related to the fact that the trigonometric system is an orthonormal uniformly bounded system.

For $1 < p \leq q \leq 2$ the case of small smoothness $\beta < r \leq 2\beta$ is settled. In this case we have the constructive bound (see [372] and [29])

$$\sigma_m(\mathbf{W}_p^r)_q \asymp m^{-(r-\beta)}. \quad (7.13)$$

Moreover, if $1 < p \leq 2 < q < \infty$ and $\beta = 1/p - 1/q < r \leq 1/p - q'/(2q)$ Theorem 7.25 below implies

$$\sigma_m(\mathbf{W}_p^r)_q \asymp m^{-(r-\beta)q/2}. \quad (7.14)$$

This (non-constructive) result actually goes back to Belinskii [29] for the larger range $\beta < r \leq q'\beta$. However, the paper [29] is hard to access. In the case of small smoothness the relations (7.13) and (7.14) show an interesting and important phenomenon. The asymptotic rate of decay of $\sigma_m(\mathbf{W}_p^r)_q$ does not depend on dimension d . The dependence on d is hidden in the constants. See also (7.10) above for a similar phenomenon in the \mathbf{B} -classes.

We note that in the case $q > 2$ Theorem 7.12 is proved in [358]. However, the proof there is not constructive – it uses a non-constructive result from [75]. In [371] we provided a constructive proof, which is based on greedy algorithms. Also, this proof works under weaker conditions on r : $r > 1/p$ instead of $r > 1/p + \eta$ for $1 < p \leq 2$. In [371] we concentrated on the case $q \geq 2$.

Let us continue with sparse trigonometric approximation in case $q = \infty$. In this case the results are not as complete as those for $1 < q < \infty$. We give here one result from [371] (the reference also contains a historical discussion).

Theorem 7.13. *We have*

$$\sigma_m(\mathbf{W}_p^r)_\infty \lesssim \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)+1/2}, & 1 < p \leq 2, \quad r > 1/p, \\ m^{-r}(\log m)^{r(d-1)+1/2}, & 2 \leq p < \infty, \quad r > 1/2. \end{cases}$$

We used recently developed techniques on greedy approximation in Banach spaces to prove Theorems 7.12 and 7.13. It is important that greedy approximation allows us not only to prove the above theorems but also to provide a constructive way for building the corresponding m -term approximants. We call this algorithm *constructive* because it provides an explicit construction with feasible one parameter optimization steps. We stress that in the setting of approximation in an infinite dimensional Banach space, which is considered in our paper, the use of term *algorithm* requires some explanation. In this paper we discuss only theoretical aspects of the efficiency (accuracy) of m -term approximation and possible ways to realize this efficiency. The *greedy algorithms* give a procedure to construct an approximant, which turns out to be a good approximant. The procedure of constructing a greedy approximant is not a numerical algorithm ready for computational implementation. Therefore, it would be more precise to call this procedure a *theoretical greedy algorithm* or *stepwise optimizing process*. Keeping this remark in mind we, however, use the term *greedy algorithm* in this paper because it has been used in previous papers and has become a standard name for procedures alike Weak Chebyshev Greedy Algorithm (see below) and for more general procedures of this type (see for instance [69], [368]). Also, the theoretical algorithms, which we use, become algorithms in a strict sense if instead of an infinite dimensional setting we consider a finite dimensional setting, replacing, for instance, the L_p space by its restriction on the set of trigonometric polynomials. We give a precise formulation from [371].

Theorem 7.14. *For $q \in (1, \infty)$ and $\mu > 0$ there exist constructive methods $A_m(f, q, \mu)$, which provide for $f \in \mathbf{W}_p^r$ an m -term approximation such that*

$$\begin{aligned} & \|f - A_m(f, q, \mu)\|_q \\ & \lesssim \begin{cases} m^{-r+\beta}(\log m)^{(d-1)(r-2\beta)}, & 1 < p \leq q \leq 2, \quad r > 2\beta + \mu, \\ m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)}, & 1 < p \leq 2 \leq q < \infty, \quad r > 1/p + \mu, \\ m^{-r}(\log m)^{r(d-1)}, & 2 \leq p \leq q < \infty, \quad r > 1/2 + \mu. \end{cases} \end{aligned}$$

Similar modification of Theorem 7.13 holds for $q = \infty$. We do not have matching lower bounds for the upper bounds in Theorem 7.13 in the case of approximation in the uniform norm L_∞ .

We now demonstrate some important features of new techniques used in the proof of the above Theorem 7.14. First of all, constructive methods $A_m(f, q, \mu)$ are built on the base of greedy-type algorithms. The reader can find a comprehensive study of greedy algorithms in [368]. One example of greedy-type algorithm – the Thresholding Greedy Algorithm – is discussed in Subsection 7.2 above. Another example – the Weak Chebyshev Greedy Algorithm – is discussed at the end of this subsection. Both of these algorithms are used in building $A_m(f, q, \mu)$. Second, it is known that the A -norm with respect to a given dictionary plays an important role in greedy approximation. In the case of the trigonometric system \mathcal{T}^d the A -norm is defined as follows

$$\|f\|_A := \sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|.$$

The following theorem and lemma from [371] play the key role in proving Theorem 7.14. Denote $\bar{m} := \max\{m, 1\}$.

Theorem 7.15. *There exist constructive greedy-type approximation methods $G_m^q(\cdot)$, which provide m -term polynomials with respect to \mathcal{T}^d with the following properties: for $2 \leq q < \infty$*

$$\|f - G_m^q(f)\|_q \leq C_1(d)(\bar{m})^{-1/2} q^{1/2} \|f\|_A, \quad \|G_m^q(f)\|_A \leq C_2(d) \|f\|_A, \quad (7.15)$$

and for $q = \infty$, $f \in \mathcal{T}(\mathbf{N}, d)$

$$\|f - G_m^\infty(f)\|_\infty \leq C_3(d)(\bar{m})^{-1/2} (\ln \vartheta(\mathbf{N}))^{1/2} \|f\|_A, \quad \|G_m^\infty(f)\|_A \leq C_4(d) \|f\|_A. \quad (7.16)$$

Taking into account importance of the A -norm for greedy approximation we prove some error bounds for approximation of classes $\mathbf{W}_A^{a,b}$ (see the definition in Lemma 7.16), which are in a style of classes \mathbf{W}_p^r with parameter a being similar to parameter r and parameter b being the one controlling the logarithmic type smoothness.

Lemma 7.16. *Define for $f \in L_1$*

$$f_l := \sum_{|\mathbf{s}|_1=l} \delta_{\mathbf{s}}(f), \quad l \in \mathbb{N}_0, \quad \mathbb{N}_0 := \mathbb{N} \cup \{0\}.$$

Consider the class

$$\mathbf{W}_A^{a,b} := \{f : \|f_l\|_A \leq 2^{-al} (\bar{l})^{(d-1)b}\}.$$

Then for $2 \leq q \leq \infty$ and $0 < \mu < a$ there is a constructive method $A_m(\cdot, q, \mu)$ based on greedy algorithms, which provides the bound for $f \in \mathbf{W}_A^{a,b}$

$$\begin{aligned} \|f - A_m(f, q, \mu)\|_q &\lesssim m^{-a-1/2} (\log m)^{(d-1)(a+b)}, \quad 2 \leq q < \infty, \\ \|f - A_m(f, \infty, \mu)\|_\infty &\lesssim m^{-a-1/2} (\log m)^{(d-1)(a+b)+1/2}. \end{aligned} \quad (7.17)$$

Proof. We prove the lemma for $m \asymp 2^n n^{d-1}$, $n \in \mathbb{N}$. Let $f \in \mathbf{W}_A^{a,b}$. We approximate f_l in L_q . By Theorem 7.15 we obtain for $q \in [2, \infty)$

$$\|f_l - G_{m_l}^q(f_l)\|_q \lesssim (\bar{m}_l)^{-1/2} \|f_l\|_A \lesssim (\bar{m}_l)^{-1/2} 2^{-al} l^{(d-1)b}. \quad (7.18)$$

We take $\mu \in (0, a)$ and specify

$$m_l := [2^{n-\mu(l-n)} l^{d-1}], \quad l = n, n+1, \dots$$

In addition we include in the approximant

$$S_n(f) := \sum_{|\mathbf{s}|_1 \leq n} \delta_{\mathbf{s}}(f).$$

Define

$$A_m(f, q, \mu) := S_n(f) + \sum_{l \geq n} G_{m_l}^q(f_l).$$

Then, we have built an m -term approximant of f with

$$m \lesssim 2^n n^{d-1} + \sum_{l \geq n} m_l \lesssim 2^n n^{d-1}.$$

The error of this approximation in L_q is bounded from above by

$$\begin{aligned} \|f - A_m(f, q, \mu)\|_q &\leq \sum_{l \geq n} \|f_l - G_{m_l}^q(f_l)\|_q \lesssim \sum_{l \geq n} (\bar{m}_l)^{-1/2} 2^{-al} l^{(d-1)b} \\ &\lesssim \sum_{l \geq n} 2^{-1/2(n-\mu(l-n))} l^{-(d-1)/2} 2^{-al} l^{(d-1)b} \lesssim 2^{-n(a+1/2)} n^{(d-1)(b-1/2)}. \end{aligned}$$

This completes the proof of lemma in the case $2 \leq q < \infty$.

Let us discuss the case $q = \infty$. The proof repeats the proof in the above case $q < \infty$ with the following change. Instead of using (7.15) for estimating an m_l -term approximation of f_l in L_q we use (7.16) to estimate an m_l -term approximation of f_l in L_∞ . Then bound (7.18) is replaced by

$$\|f_l - G_{m_l}^\infty(f_l)\|_\infty \lesssim (\bar{m}_l)^{-1/2} (\ln 2^l)^{1/2} \|f_l\|_A \lesssim (\bar{m}_l)^{-1/2} l^{1/2} 2^{-al} l^{(d-1)b}. \quad (7.19)$$

The extra factor $l^{1/2}$ in (7.19) gives an extra factor $(\log m)^{1/2}$ in (7.17). ■

We now consider the case $\sigma_m(\mathbf{W}_1^r)_p$, which is not covered by Theorem 7.12. The function $F_r(\mathbf{x})$ belongs to the closure in L_q of \mathbf{W}_1^r , $r > 1 - 1/q$, and, therefore, on the one hand

$$\sigma_m(\mathbf{W}_1^r)_q \geq \sigma_m(F_r(\mathbf{x}))_q.$$

On the other hand, it follows from the definition of \mathbf{W}_1^r that for any $f \in \mathbf{W}_1^r$

$$\sigma_m(f)_q \leq \sigma_m(F_r(\mathbf{x}))_q.$$

Thus,

$$\sigma_m(\mathbf{W}_1^r)_q = \sigma_m(F_r(\mathbf{x}))_q.$$

The following results on $\sigma_m(F_r(\mathbf{x}))_q$ are from [371].

Theorem 7.17. *We have*

$$\sigma_m(F_r(\mathbf{x}))_q \asymp \begin{cases} m^{-r+1-1/q} (\log m)^{(d-1)(r-1+2/q)}, & 1 < q \leq 2, \quad r > 1 - 1/q, \\ m^{-r+1/2} (\log m)^{r(d-1)}, & 2 \leq q < \infty, \quad r > 1. \end{cases}$$

The upper bounds are provided by a constructive method $A_m(\cdot, q, \mu)$ based on greedy algorithms.

Theorem 7.18. *We have*

$$\sigma_m(F_r(\mathbf{x}))_\infty \lesssim m^{-r+1/2} (\log m)^{r(d-1)+1/2}, \quad r > 1.$$

The bounds are provided by a constructive method $A_m(\cdot, \infty, \mu)$ based on greedy algorithms.

We now proceed to classes \mathbf{H}_p^r and $\mathbf{B}_{p,\theta}^r$. The following theorem was proved in [336] (see also [340], Ch.4). The proofs from [336] and [340] are constructive.

Theorem 7.19. *Let $1 < p \leq q \leq 2$, $r > \beta$. Then*

$$\sigma_m(\mathbf{H}_p^r)_q \asymp m^{-r+\beta}(\log m)^{(d-1)(r-\beta+1/q)}.$$

The following analog of Theorem 7.12 for classes \mathbf{H}_p^r was proved in [271]. The proof in [271] in the case $q > 2$ is not constructive.

Theorem 7.20. *One has*

$$\sigma_m(\mathbf{H}_p^r)_q \asymp \begin{cases} m^{-r+\beta}(\log m)^{(d-1)(r-\beta+1/q)}, & 1 < p \leq q \leq 2, \quad r > \beta, \\ m^{-r+\eta}(\log m)^{(d-1)(r-1/p+1)}, & 1 < p \leq 2 \leq q < \infty, \quad r > 1/p, \\ m^{-r}(\log m)^{(d-1)(r+1/2)}, & 2 \leq p \leq q < \infty, \quad r > 1/2. \end{cases}$$

The third line of Theorem 7.20 combined with (7.2) shows that the trigonometric system is an optimal in the sense of order orthonormal system for the m -term approximation of the classes \mathbf{H}_p^r in L_q for $2 \leq p, q < \infty$. The following proposition is from [371].

Proposition 7.21. *The upper bounds in Theorem 7.20 are provided by a constructive method $A_m(\cdot, q, \mu)$ based on greedy algorithms.*

In the case $q = \infty$ we have (see [371]).

Theorem 7.22. *We have*

$$\sigma_m(\mathbf{H}_p^r)_\infty \lesssim \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-1/p+1)+1/2}, & 1 < p \leq 2, \quad r > 1/p, \\ m^{-r}(\log m)^{(r+1/2)(d-1)+1/2}, & 2 \leq p < \infty, \quad r > 1/2. \end{cases}$$

The upper bounds are provided by a constructive method $A_m(\cdot, \infty, \mu)$ based on greedy algorithms.

For a non-constructive proof of the bound of Theorem 7.22 in the case $2 \leq p < \infty$ see [30].

We now proceed to classes $\mathbf{B}_{p,\theta}^r$. It will be convenient for us to use the following slight modification of classes $\mathbf{B}_{p,\theta}^r$. Define

$$\|f\|_{\mathbf{H}_{p,\theta}^r} := \sup_n \left(\sum_{\mathbf{s}: |\mathbf{s}|_1 = n} \left(\|\delta_{\mathbf{s}}(f)\|_p 2^{r|\mathbf{s}|_1} \right)^\theta \right)^{1/\theta}$$

and

$$\mathbf{H}_{p,\theta}^r := \{f : \|f\|_{\mathbf{H}_{p,\theta}^r} \leq 1\}.$$

The best m -term approximations of classes $\mathbf{B}_{p,\theta}^r$ are studied in detail by A.S. Romanyuk [271]. There is the following extension of Theorem 7.20 (see [271]).

Theorem 7.23. *One has*

$$\sigma_m(\mathbf{B}_{p,\theta}^r)_q \asymp m^{-r+\beta}(\log m)^{(d-1)(r-\beta+1/p-1/\theta)+}, \quad 1 < p \leq q \leq 2, \quad r > \beta,$$

and

$$\begin{aligned} \sigma_m(\mathbf{B}_{p,\theta}^r)_q &\asymp \sigma_m(\mathbf{H}_{p,\theta}^r)_q \\ &\asymp \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-1/p+1-1/\theta)}, & 1 < p \leq 2 \leq q < \infty, \quad r > 1/p, \\ m^{-r}(\log m)^{(d-1)(r+1/2-1/\theta)}, & 2 \leq p \leq q < \infty, \quad r > 1/2. \end{cases} \end{aligned} \quad (7.20)$$

The following proposition is from [371].

Proposition 7.24. *The upper bounds in Theorem 7.23 are provided by a constructive method based on greedy algorithms.*

In the case of small smoothness the following result is known (see [271, Thm. 2.1] engl. version).

Theorem 7.25. *Let $1 \leq p \leq 2 < q < \infty$ and $1 \leq \theta \leq \infty$. Then*

$$\sigma_m(\mathbf{B}_{p,\theta}^r)_q \asymp \begin{cases} m^{-1/2}(\log m)^{d(1-1/\theta)}, & r = 1/p, \\ m^{-q(r-\beta)/2}(\log m)^{(d-1)(q-1)(r-1/p+q'/(q\theta'))_+}, & \beta < r < 1/p. \end{cases}$$

The case $1 < p \leq 2 < q < \infty$, $\theta = \infty$, $\beta < r < 1/p$ was considered by Temlyakov in [372, Thm. 3.4]. There the upper bounds were achieved by a constructive greedy-type algorithm (compared to the proof of [271, Thm. 2.1], non-constructive upper bounds). Theorem 7.25 is also true in more general settings, see the recent paper by Stasyuk [309].

We formulate some known results in the case $1 < q \leq p \leq \infty$.

Theorem 7.26. *Let $1 < q \leq p < \infty$, $r > 0$. Then*

$$\sigma_m(\mathbf{W}_p^r)_q \asymp m^{-r}(\log m)^{(d-1)r}.$$

The upper bound in Theorem 7.26 follows from error bounds for approximation by the hyperbolic cross polynomials (see [340], Ch.2, §2)

$$E_{Q_n}(\mathbf{W}_q^r, L_q) \lesssim 2^{-rn}, \quad 1 < q < \infty.$$

The lower bound in Theorem 7.26 was proved in [195].

The following result for \mathbf{H}_p^r classes is known.

Theorem 7.27. *Let $q \leq p$, $2 \leq p \leq \infty$, $1 < q < \infty$, $r > 0$. Then*

$$\sigma_m(\mathbf{H}_p^r)_q \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

The lower bound for all $q > 1$

$$\sigma_m(\mathbf{H}_\infty^r)_q \gtrsim m^{-r}(\log m)^{(d-1)(r+1/2)}$$

was obtained in [195]. The matching upper bounds follow from approximation by the hyperbolic cross polynomials (see [340], Ch.2, Theorem 2.2)

$$E_{Q_n}(\mathbf{H}_q^r)_q := \sup_{f \in \mathbf{H}_q^r} E_{Q_n}(f)_q \asymp n^{(d-1)/2} 2^{-rn}, \quad 2 \leq q < \infty.$$

The following result for \mathbf{B} classes was proved in [271].

Theorem 7.28. *Let $1 < q \leq p < \infty$, $2 \leq p < \infty$, $1 < q < \infty$, $r > 0$. Then*

$$\sigma_m(\mathbf{B}_{p,\theta}^r)_q \asymp m^{-r}(\log m)^{(d-1)(r+1/2-1/\theta)_+}.$$

We are interested in the following fundamental problem of sparse approximation.

Problem. How to design a practical algorithm that builds sparse approximations comparable to best m -term approximations?

It was demonstrated in the paper [370] that the Weak Chebyshev Greedy Algorithm (WCGA), which we define momentarily, is a solution to the above problem for a special class of dictionaries.

Let X be a real Banach space with norm $\|\cdot\| := \|\cdot\|_X$. We say that a set of elements (functions) \mathcal{D} from X is a dictionary if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$), and the closure of $\text{span } \mathcal{D}$ is X . For a nonzero element $g \in X$ we let F_g denote a norming (peak) functional for g :

$$\|F_g\|_{X^*} = 1, \quad F_g(g) = \|g\|_X.$$

The existence of such a functional is guaranteed by the Hahn-Banach theorem.

Let $t \in (0, 1]$ be a given weakness parameter. Define the Weak Chebyshev Greedy Algorithm (WCGA) (see [362]) as a generalization for Banach spaces of the Weak Orthogonal Matching Pursuit (WOMP). In a Hilbert space the WCGA coincides with the WOMP. The WOMP is very popular in signal processing, in particular, in compressed sensing. In case $t = 1$, WOMP is called Orthogonal Matching Pursuit (OMP).

Weak Chebyshev Greedy Algorithm (WCGA). Let f_0 be given. Then for each $m \geq 1$ we have the following inductive definition.

- (1) $\varphi_m := \varphi_m^{c,t} \in \mathcal{D}$ is any element satisfying

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|.$$

- (2) Define $\Phi_m := \Phi_m^t := \text{span} \{\varphi_j\}_{j=1}^m$, and define $G_m := G_m^{c,t}$ to be the best approximant to f_0 from Φ_m .

- (3) Let $f_m := f_m^{c,t} := f_0 - G_m$.

The trigonometric system is a classical system that is known to be difficult to study. In [370] we study among other problems the problem of nonlinear sparse approximation with respect to it. Let \mathcal{RT} denote the real trigonometric system $1, \sin x, \cos x, \dots$ on $[0, 2\pi]$ and let \mathcal{RT}_p to be its version normalized in $L_p([0, 2\pi])$. Denote $\mathcal{RT}_p^d := \mathcal{RT}_p \times \dots \times \mathcal{RT}_p$ the d -variate trigonometric system. We need to consider the real trigonometric system because the algorithm WCGA is well studied for the real Banach space. In order to illustrate performance of the WCGA we discuss in this section the above mentioned problem for the trigonometric system. The following Lebesgue-type inequality for the WCGA was proved in [370].

Theorem 7.29. *Let \mathcal{D} be the normalized in L_p , $2 \leq p < \infty$, real d -variate trigonometric system. Then for any $f_0 \in L_p$ the WCGA with weakness parameter t gives*

$$\|f_{C(t,p,d)m \ln(m+1)}\|_p \leq C \sigma_m(f_0, \mathcal{D})_p. \quad (7.21)$$

The Open Problem 7.1 (p. 91) from [364] asks if (7.21) holds without an extra $\ln(m+1)$ factor. Theorem 7.29 is the first result on the Lebesgue-type inequalities for the WCGA with respect to the trigonometric system. It provides a progress in solving the above mentioned open problem, but the problem is still open.

Theorem 7.29 shows that the WCGA is very well designed for the trigonometric system. It was shown in [370] that an analog of (7.21) holds for uniformly bounded orthogonal systems.

As a direct corollary of Theorems 7.29 and 7.12 we obtain the following result (see [371]).

Theorem 7.30. *Let $q \in [2, \infty)$. Apply the WCGA with weakness parameter $t \in (0, 1]$ to $f \in L_q$ with respect to the real trigonometric system \mathcal{RT}_q^d . If $f \in \mathbf{W}_p^r$, then we have*

$$\|f_m\|_q \lesssim \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)+r-\eta}, & 1 < p \leq 2, \quad r > 1/p, \\ m^{-r}(\log m)^{rd}, & 2 \leq p < \infty, \quad r > 1/2. \end{cases}$$

7.6 Open problems

It is well known that the extreme cases, when one of the parameters p or q takes values 1 or ∞ are difficult in the hyperbolic cross approximation theory. Often, study of these cases requires special techniques. Many of the problems, which involve the extreme values of parameters, are still open. Also the case of small smoothness is still open in many settings.

Open problem 7.1. Find a constructive method, which provides the order of $\sigma_m(\mathbf{W}_p^r)_q$, $2 \leq p \leq q < \infty$, $\beta < r \leq 1/2$.

Open problem 7.2. Find the order of $\sigma_m(\mathbf{W}_p^r)_\infty$, $1 \leq p \leq \infty$, $r > 1/p$.

Open problem 7.3. Find the order of $\sigma_m(\mathbf{W}_p^r)_1$, $1 \leq p \leq \infty$, $r > 0$.

Open problem 7.4. Find the order of $\sigma_m(\mathbf{W}_\infty^r)_q$, $1 \leq q \leq \infty$, $r > 0$.

We formulated the above problems for the \mathbf{W} classes. Those problems are open for the \mathbf{H} and \mathbf{B} classes as well. In addition the following problem is open for the \mathbf{H} and \mathbf{B} classes.

Open problem 7.5. Find the order of $\sigma_m(\mathbf{H}_p^r)_q$ and $\sigma_m(\mathbf{B}_{p,\theta}^r)_q$ for $1 \leq q < p \leq 2$, $r > 0$.

8 Numerical integration

8.1 The problem setting

A cubature rule $\Lambda_m(f, \xi)$ approximates the integral $I(f) = \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x}$ by computing a weighted sum of finitely many function values at $X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\}$, i.e.,

$$\Lambda_m(f, X_m) := \sum_{i=1}^m \lambda_i f(\mathbf{x}^i), \quad (8.1)$$

where the d -variate function f is assumed to belong to some (quasi-)normed function space $\mathbf{F} \subset C([0, 1]^d)$. The optimal error with respect to the class \mathbf{F} is given by

$$\kappa_m(\mathbf{F}) := \inf_{X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\} \subset [0, 1]^d} \inf_{\lambda_1, \dots, \lambda_m \in \mathbb{R}} \Lambda_m(\mathbf{F}, X_m), \quad (8.2)$$

where

$$\Lambda_m(\mathbf{F}, X_m) := \sup_{f \in \mathbf{F}} \left| I(f) - \sum_{i=1}^m \lambda_i f(\mathbf{x}^i) \right|.$$

It is clear that $f \in \mathbf{F}$ has to be defined at the nodes $\{\mathbf{x}^1, \dots, \mathbf{x}^m\}$. For that reason we always assume $\mathbf{F} \hookrightarrow C(\mathbb{T}^d)$. For simplicity, in contrast to the other sections, we represent the d -torus here as $\mathbb{T}^d = [0, 1]^d$. Our main interest in this section is to present known optimal results (in the sense of order) in the number of nodes m for numerical integration of classes of functions with bounded mixed smoothness.

8.2 Lower bounds

The lower bounds that we want to present are valid for *arbitrary* cubature formulas. For this we study the quantity $\kappa_m(\mathbf{F})$ for the spaces $\mathbf{B}_{p,\theta}^r$ and \mathbf{W}_p^r . We will provide two approaches for “fooling” the given cubature formula in order to obtain asymptotically sharp lower bounds.

Fooling polynomials

Let us start with the following approach which has been a big breakthrough at that time. The theorem is from [345]. The idea is to construct “bad” trigonometric polynomials which fool the given cubature formula in the following way.

Theorem 8.1. *Let $1 \leq p < \infty$ and $r > 1/p$. Then we have*

$$\kappa_m(\mathbf{W}_p^r) \gtrsim m^{-r} (\log m)^{(d-1)/2}.$$

The proof of Theorem 8.1 is based on the existence of special trigonometric polynomials given by Theorem 2.3 above. In fact, Theorem 2.3 is used to prove the following assertion.

Lemma 8.2. *Let the coordinates of the vector \mathbf{s} be natural numbers and $|\mathbf{s}|_1 = n$. Then for any $N \leq 2^{n-1}$ and an arbitrary cubature formula $\Lambda_N(\cdot, X_N)$ with N nodes there is a $t_{\mathbf{s}} \in \mathcal{T}(2^{\mathbf{s}}, d)$ such that $\|t_{\mathbf{s}}\|_{\infty} \leq 1$ and*

$$\hat{t}_{\mathbf{s}}(0) - \Lambda_N(t_{\mathbf{s}}, X_N) \geq C(d) > 0.$$

Now we choose for a given m a number n such that

$$m \leq 2^{n-1} < 2m.$$

Consider the polynomial

$$t(\mathbf{x}) = \sum_{|\mathbf{s}|_1=n} t_{\mathbf{s}}(\mathbf{x}),$$

where $t_{\mathbf{s}}$ are polynomials from Lemma 8.2 with $N = m$. Then

$$\hat{t}(0) - \Lambda_m(t, X_m) \geq C(d)n^{d-1}.$$

The proof of Theorem 8.1 was completed by establishing that if $2 \leq p < \infty$ then

$$\|t\|_{\mathbf{W}_p^r} \lesssim \|t\|_{\mathbf{B}_{p,2}^r} \lesssim 2^{rn} n^{(d-1)/2}.$$

Theorem 8.1 gives the same lower bound for different parameters $1 \leq p < \infty$. It is clear that the bigger the p the stronger the statement. It was pointed out in [371] that the above example also provides the lower bound for the Besov classes.

Theorem 8.3. *Let $1 \leq p, \theta \leq \infty$ and $r > 1/p$. We have the following lower bound for the Besov classes $\mathbf{B}_{p,\theta}^r$*

$$\kappa_m(\mathbf{B}_{p,\theta}^r) \gtrsim m^{-r} (\log m)^{(d-1)(1-1/\theta)}, \quad 1 \leq p \leq \infty, \quad 1 \leq \theta \leq \infty.$$

Indeed, it is easy to check that

$$\|t\|_{\mathbf{B}_{p,\theta}^r} \lesssim 2^{rn} n^{(d-1)/\theta}.$$

The following lower bounds for numerical integration with respect to a special class of nodes were obtained recently in [371]. Let $\mathbf{s} = (s_1, \dots, s_d)$, $s_j \in \mathbb{N}_0$, $j = 1, \dots, d$. We associate with \mathbf{s} a $W(\mathbf{s})$ as follows: denote

$$w(\mathbf{s}, \mathbf{x}) := \prod_{j=1}^d \sin(2\pi 2^{s_j} x_j)$$

and define

$$W(\mathbf{s}) := \{\mathbf{x} : w(\mathbf{s}, \mathbf{x}) = 0\}.$$

Definition 8.4. *We say that a set of nodes $X_m := \{\mathbf{x}^i\}_{i=1}^m$ is an (n, ℓ) -net if $|X_m \setminus W(\mathbf{s})| \leq 2^\ell$ for all \mathbf{s} such that $|\mathbf{s}|_1 = n$.*

Theorem 8.5. *For any cubature formula $\Lambda_m(\cdot, X_m)$ with respect to a $(n, n-1)$ -net X_m we have for $1 \leq p < \infty$ that*

$$\Lambda_m(\mathbf{W}_p^r, X_m) \gtrsim 2^{-rn} n^{(d-1)/2}, \quad m \in \mathbb{N}.$$

In the same way as a slight modification of the proof of Theorem 8.1 gave Theorem 8.3 a similar modification of the proof of Theorem 8.5 gives the following result.

Theorem 8.6. *Let $1 \leq p, \theta \leq \infty$ and $r > 1/p$. Then for any cubature formula $\Lambda_m(\cdot, X_m)$ with respect to a $(n, n-1)$ -net X_m we have that*

$$\Lambda_m(\mathbf{B}_{p,\theta}^r, X_m) \gtrsim 2^{-rn} n^{(d-1)(1-1/\theta)}, \quad n \in \mathbb{N}.$$

In case $\mathbf{W}_{1,0}^r$ with $r > 1$ things can be “improved”. We obtain a larger lower bound than the one in Theorem 8.5 by slightly shrinking the class of cubature formulas via the following assumption

$$\sum_{\mu=1}^m |\lambda_\mu| \leq B.$$

The corresponding minimal error with respect to a class \mathbf{F} is then defined by

$$\kappa_m^B(\mathbf{F}) := \inf_{\sum_{\mu=1}^m |\lambda_\mu| \leq B} \inf_{X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\}} \Lambda_m(\mathbf{F}, X_m).$$

The following result is from [353].

Theorem 8.7. *Let $r > 1$. Then there is a constant $C = C(r, B, d)$ such that*

$$\kappa_m^B(\mathbf{W}_{1,0}^r) \geq C(r, B, d) m^{-r} (\log m)^{d-1}.$$

Local fooling functions

The idea of using fooling functions to prove lower bounds for asymptotic characteristics of functions goes back to Bakhvalov [18]. The following approach relies on atomic decompositions, a modern tool in function space theory, to control the norms of superpositions of local bumps, see Section 5.5. This powerful approach allows as well to treat the quasi-Banach situation where $p < 1$, see for instance [388]. However, what concerns this survey we will restrict ourselves to the Banach space case. With the results from Proposition 5.25 we are in a position to define test functions of type (5.47) in order to prove the required lower bounds. By (5.48) we are able to control the norm $\|\cdot\|_{\mathbf{B}_{p,\theta}^r}$. Following [109, Thm. 4.1] and [388] we will use test functions of type

$$g_{r,\theta} := C 2^{-r\ell} \ell^{-(d-1)/\theta} \sum_{|\mathbf{j}|_1 = \ell+1} \sum_{\mathbf{k} \in K_{\mathbf{j}}(X_m)} a_{\mathbf{j},\mathbf{k}}, \quad (8.3)$$

where $K_{\mathbf{j}}(X_m) \subset \{0, \dots, 2^{j_1-1} \times \dots \times \{0, \dots, 2^{j_d-1}\}$ depends on the set of integration nodes $X_m := \{\mathbf{x}^1, \dots, \mathbf{x}^m\}$ with $m = 2^\ell$.

Let us now give a different proof of Theorems 8.1 and 8.3. One only needs to prove Theorem 8.3. In fact, Theorem 8.1 follows from Theorem 8.3 together with the embeddings $\mathbf{B}_{2,2}^r \hookrightarrow \mathbf{W}_p^r$ if $p \leq 2$ and $\mathbf{B}_{p,2}^r \hookrightarrow \mathbf{W}_p^r$ if $p > 2$. Following the arguments in [109, Thm. 4.1] let m be given and $X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\} \subset [0, 1]^d$ be an arbitrary set of $m = 2^\ell$ points. Since $Q_{\mathbf{j},\mathbf{k}} \cap Q_{\mathbf{j},\mathbf{k}'} = \emptyset$ for $\mathbf{k} \neq \mathbf{k}'$ we have for every $|\mathbf{j}|_1 = \ell + 1$ a set of 2^ℓ cubes of the form $2^{-\mathbf{j}}\mathbf{k} + 2^{-\mathbf{j}}Q$ which do not intersect the nodes X_m . We choose the test function (8.3) where $K_{\mathbf{j}}(X_m)$ is the set of those \mathbf{k} referring to those cubes. By (5.48) the $\mathbf{B}_{p,\theta}^r$ norm of those functions is uniformly bounded (in ℓ). The result in case $p, \theta \geq 1$ follows from the observation

$$\int_{[0,1]^d} g_{r,\theta} d\mathbf{x} \asymp 2^{-\ell r} \ell^{(d-1)(1-1/\theta)}.$$

Of course, a cubature rule admitted in (8.2), which uses the points X_m , produces a zero output. For parameters $\min\{p, \theta\} < 1$ the fooling function (8.3) has to be slightly modified.

8.3 Cubature on Smolyak grids

Here we will discuss cubature on Smolyak grids. The Smolyak grid of level ℓ is given by

$$SG^d(\ell) := \bigcup_{k_1 + \dots + k_d \leq \ell} I_{k_1} \times \dots \times I_{k_d} \quad (8.4)$$

where $I_k := \{2^{-k}n : n = 0, \dots, 2^k - 1\}$. We consider the cubature formulas $\Lambda_m(f, SG^d(\ell))$ on Smolyak grids $SG^d(\ell)$ given by

$$\Lambda_m(f, SG^d(\ell)) = \sum_{\mathbf{x}^i \in SG^d(\ell)} \lambda_i f(\mathbf{x}^i). \quad (8.5)$$

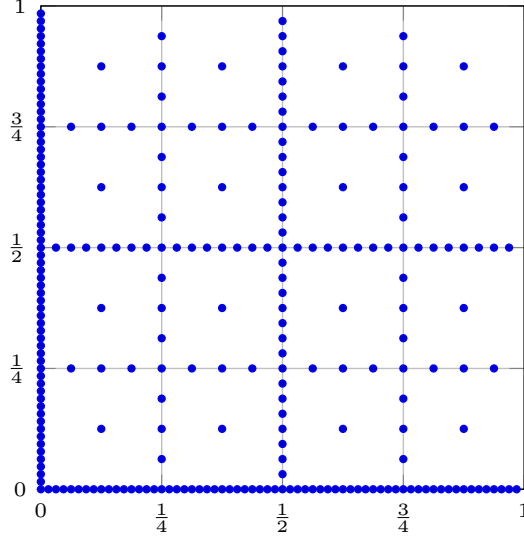


Figure 15: A sparse grid in $d = 2$ with $N = 256$ points

Here we have a degree of freedom when choosing the weights $\{\lambda_i\}_{\mathbf{x}^i \in G^d(\ell)}$. However, it turns out that any cubature formula on Smolyak grids behaves asymptotically worse than the optimal rules discussed below. The following theorem presents the approach from [109]. It relies on the same principle as above of using superpositions of local bump functions to fool the algorithm. Taking into account that a sparse grid of level ℓ contains $m \asymp 2^\ell \ell^{d-1}$ points we obtain the following general result.

Theorem 8.8. *Let $1 \leq p, \theta \leq \infty$ and $r > 1/p$.*

(i) *Then*

$$\begin{aligned} \Lambda_m(\mathbf{B}_{p,\theta}^r, SG^d(\ell)) &\gtrsim 2^{-\ell r} \ell^{(d-1)(1-1/\theta)} \\ &\asymp m^{-r} (\log m)^{(d-1)(r+1-1/\theta)}. \end{aligned} \quad (8.6)$$

(ii) *If $1 < p < \infty$ and $r > 0$ then*

$$\begin{aligned} \Lambda_m(\mathbf{W}_p^r, SG^d(\ell)) &\gtrsim 2^{-\ell r} \ell^{(d-1)/2} \\ &\asymp m^{-r} (\log m)^{(d-1)(r+1/2)}. \end{aligned} \quad (8.7)$$

Remark 8.9. *Those lower bounds show that numerical integration on Smolyak grids is “not easier” than Smolyak sampling, see Section 5 above. However, it is also not harder. Note, that the results in Section 5 show that there is a sampling algorithm of the form*

$$A_m(f, G^d(\ell)) = \sum_{\mathbf{x}^i \in G^d(\ell)} f(\mathbf{x}^i) a_{\mathbf{x}^i}(\cdot)$$

which approximates f well in $L_1(\mathbb{T}^d)$. Taking

$$\left| \int_{\mathbb{T}^d} f(\mathbf{x}) d\mathbf{x} - \int_{\mathbb{T}^d} \sum_{\mathbf{x}^i \in G^d(\ell)} f(\mathbf{x}^i) a_{\mathbf{x}^i}(\mathbf{x}) d\mathbf{x} \right| \leq \|f - A_m(f, G^d(\ell))\|_1$$

into account we obtain with $\lambda_{\mathbf{x}^i} := \int_{\mathbb{T}^d} a_{\mathbf{x}^i}(\mathbf{x}) d\mathbf{x}$ a cubature formula of type (8.5) where the error is bounded by the approximation error in $L_1(\mathbb{T}^d)$. Hence, by the results presented in Section 5 we see that the lower bounds (8.6) and (8.7) are sharp in the sense, that there is a cubature formula on the Smolyak grid with matching upper bounds.

In [371] Temlyakov observed that the same bounds can be also obtained from the approach based on the general Theorem 2.3. Let us briefly discuss how Theorems 8.5 and 8.6 imply bounds (8.6) and (8.7) (see [371]). It is easy to check that $SG^d(n) \subset W(\mathbf{s})$ with any \mathbf{s} such that $|\mathbf{s}|_1 = n$. Indeed, let $\xi(\mathbf{n}, \mathbf{k}) \in SG^d(n)$. Take any \mathbf{s} with $|\mathbf{s}|_1 = n$. Then $|\mathbf{s}|_1 = |\mathbf{n}|_1$ and there exists j such that $s_j \geq n_j$. For this j we have

$$\sin 2^{s_j} \xi(\mathbf{n}, \mathbf{k})_j = \sin 2^{s_j} \pi k_j 2^{-n_j} = 0 \quad \text{and} \quad w(\mathbf{s}, \xi(\mathbf{n}, \mathbf{k})) = 0.$$

This means that $SG^d(n)$ is an (n, l) -net for any l . We note that $|SG^d(n)| \asymp 2^n n^{d-1}$. It is known (see [335]) that there exists a cubature formula $(\Lambda, SG^d(n))$ such that

$$\Lambda(\mathbf{H}_p^r, G^d(n)) \lesssim 2^{-rn} n^{d-1}, \quad 1 \leq p \leq \infty, \quad r > 1/p. \quad (8.8)$$

Theorem 8.6 with $\theta = \infty$ shows that the bound (8.8) is sharp. Moreover, Theorem 8.6 shows that even an addition of extra 2^{n-1} arbitrary nodes to $SG^d(n)$ will not improve the bound in (8.8).

Novak and Ritter [249] studied Smolyak cubature based on the one-dimensional Clenshaw-Curtis rule with numerical experiments. Later, Gerstner and Griebel [143] investigated and compared several variants of Smolyak cubature rules based on different one-dimensional quadrature schemes like Gauss (Patterson), trapezoidal and Clenshaw-Curtis quadrature rules. Their numerical experiments show that Gauss (Patterson) performs best (among the other methods mentioned) for the considered examples. The Clenshaw-Curtis Smolyak cubature rule has been also considered from the viewpoint of tractability for a class of infinite times differentiable functions. In Hinrichs, Novak, M. Ullrich [176] the authors prove “weak tractability” in this setting based on the polynomial exactness of this cubature rule.

8.4 The Fibonacci cubature formulas

The Fibonacci cubature formulas have been studied by several authors, see for instance Bakhvalov [16], Temlyakov [347, 349, 353], Dũng, Ullrich [109] and the references in [352, Chapt. IV]. This cubature formula behaves asymptotically optimal within the bivariate function classes of interest here. It represents a rank-1-lattice rule. Note, that for higher dimensions optimal rank-1-lattice rules are not known.

Theorem 8.10. (i) Let $d = 2$. For $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$ we have

$$\kappa_m(\mathbf{W}_p^r(\mathbb{T}^2)) \asymp m^{-r}(\log m)^{1/2}.$$

(ii) For $1 \leq p, \theta \leq \infty$, and $r > 1/p$ we have

$$\kappa_m(\mathbf{B}_{p,\theta}^r(\mathbb{T}^2)) \asymp m^{-r}(\log m)^{1-1/\theta}.$$

Note that all the lower estimates are provided by the results in Subsection 8.2 above. The upper bounds in both above theorems are obtained by the use of the Fibonacci cubature formulas. The Fibonacci cubature formulas are defined as follows

$$\Phi_n(f) := b_n^{-1} \sum_{\mu=1}^{b_n} f(\mu/b_n, \{\mu b_{n-1}/b_n\}),$$

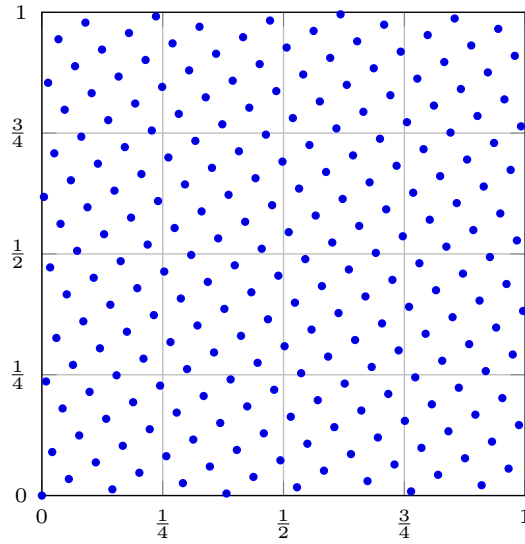


Figure 16: A Fibonacci lattice with $N = 233$ points

where $b_0 = b_1 = 1$, $b_n = b_{n-1} + b_{n-2}$ are the Fibonacci numbers and $\{x\}$ is the fractional part of the number x . Those cubature formulas are designed for periodic functions and they are exact on the trigonometric polynomials with frequencies in the hyperbolic cross $\Gamma(\gamma b_n)$, where $\gamma > 0$ is a universal constant.

For a function class $\mathbf{F}(\mathbb{T}^2)$ of bivariate functions we denote

$$\Phi_n(\mathbf{F}(\mathbb{T}^2)) := \sup_{f \in \mathbf{F}(\mathbb{T}^2)} \left| \Phi_n(f) - \int_{\mathbb{T}^2} f(\mathbf{x}) d\mathbf{x} \right|,$$

where $\mathbb{T}^2 = [0, 1]^2$ represents the 2-torus.

Proposition 8.11. (i) For $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$ we have

$$\Phi_n(\mathbf{W}_p^r(\mathbb{T}^2)) \asymp b_n^{-r}(\log b_n)^{1/2}.$$

(ii) For $1 \leq p, \theta \leq \infty$, and $r > 1/p$ we have

$$\Phi_n(\mathbf{B}_{p,\theta}^r(\mathbb{T}^2)) \asymp b_n^{-r}(\log b_n)^{1-1/\theta}.$$

Proposition 8.11 and Theorem 8.10 show that the Fibonacci cubature formulas are optimal in the sense of order in the stated situations.

Small smoothness and limiting cases

Here we are particularly interested in the limiting situations $\mathbf{B}_{p,1}^{1/p}(\mathbb{T}^2)$, $\mathbf{W}_1^r(\mathbb{T}^2)$ and the situation of “small smoothness” occurring whenever we deal with $\mathbf{W}_p^r(\mathbb{T}^2)$ with $2 < p < \infty$ and $1/p < r \leq 1/2$, which is not covered by Theorem 8.10 and Proposition 8.11.

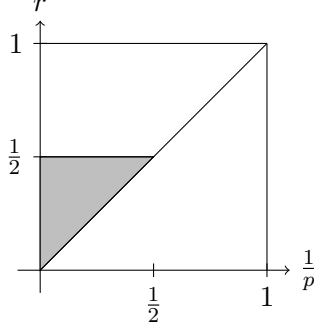


Figure 17: The region of “small smoothness”

Proposition 8.12. (i) *Let $r > 1$ then it holds*

$$\Phi_n(\mathbf{W}_{1,\alpha}^r(\mathbb{T}^2)) \asymp b_n^{-r} \log b_n.$$

(ii) *If $r > 1/2$ we have*

$$\Phi_n(\mathbf{W}_\infty^r(\mathbb{T}^2)) \asymp b_n^{-r} \sqrt{\log b_n}.$$

(iii) *If $1 \leq p < \infty$ then*

$$\Phi_n(\mathbf{B}_{p,1}^{1/p}(\mathbb{T}^2)) \asymp b_n^{-1/p}.$$

Relation (i) is proved in [353]. The lower bound in (i) for $\alpha = 0$ follows from Theorem 8.7. The upper bound in part (iii) follows directly from (3.17) in [109]. for the lower bound we refer to [388, Thm. 7.3]. Note, that the embedding $\mathbf{B}_{p,1}^{1/p} \hookrightarrow C(\mathbb{T}^d)$ holds true, see Lemma 3.12,(iii) above.

Next we state the results for small smoothness.

Proposition 8.13. *Let $2 < p < \infty$ and $1/p < r \leq 1/2$. Then*

$$\Phi_n(\mathbf{W}_p^r) \asymp \begin{cases} b_n^{-r} (\log b_n)^{1-r}, & 1/p < r < 1/2; \\ b_n^{-r} \sqrt{(\log b_n)(\log \log b_n)}, & r = 1/2. \end{cases}$$

One observes a different behavior in the log-power and an additional log log if $r = 1/2$. A similar effect seems to hold in the multivariate situation when dealing with Frolov’s cubature, see Theorem 8.19 below for the upper bounds. Note, that in contrast to the result for the Fibonacci cubature rule there are so far no sharp lower bounds for the multivariate situation.

8.5 The Frolov cubature formulas

The Frolov cubature formulas were introduced and studied in [128, 129]. The reader can find a detailed discussion of this topic in [352, Chapt. IV], [365], and [386]. For the analysis of this method in several Besov and Triebel-Lizorkin type spaces of mixed smoothness we refer to Dubinin [114, 115] and the recent papers [388, 239]. The Frolov cubature formulas are used in the proof of the upper bounds in the following theorem.

Theorem 8.14. (i) For $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$ we have

$$\kappa_m(\mathbf{W}_p^r) \asymp m^{-r}(\log m)^{(d-1)/2}.$$

(ii) For $1 \leq p, \theta \leq \infty$, and $r > 1/p$ we have

$$\kappa_m(\mathbf{B}_{p,\theta}^r) \asymp m^{-r}(\log m)^{(d-1)(1-1/\theta)}.$$

(iii) For $1 \leq p < \infty$ we have

$$\kappa_m(\mathbf{B}_{p,1}^{1/p}) \asymp m^{-1/p}.$$

The lower bounds in (i) are provided by Theorem 8.1 and in (ii) by Theorem 8.3. The lower bound in (iii) is proven in [388, Thm. 7.3].

Contrary to the case of the Fibonacci (see above) and the Korobov (see below) cubature formulas the Frolov cubature formulas defined below are not designed for a direct application to periodic functions. As a result one needs to use a two step strategy. We begin with a definition. The following results, see for instance [352, IV.4] and [186], play a fundamental role in the construction of these formulas.

Theorem 8.15. *There exists a matrix A such that the lattice $L(\mathbf{m}) = A\mathbf{m}$, i.e.,*

$$L(\mathbf{m}) := \begin{pmatrix} L_1(\mathbf{m}) \\ \vdots \\ L_d(\mathbf{m}) \end{pmatrix},$$

where \mathbf{m} is a (column) vector with integer coordinates, has the following properties

- (1) $\left| \prod_{j=1}^d L_j(\mathbf{m}) \right| \geq 1$ for all $\mathbf{m} \neq \mathbf{0}$;
- (2) each parallelepiped P with volume $|P|$ whose edges are parallel to the coordinate axes contains no more than $|P| + 1$ lattice points.

There is a constructive approach in choosing the lattice generating matrix A . In the original paper by Frolov [128] a Vandermonde matrix

$$A = \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^{d-1} \\ 1 & \xi_2 & \cdots & \xi_2^{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_d & \cdots & \xi_d^{d-1} \end{pmatrix} \quad (8.9)$$

has been considered, where ξ_1, \dots, ξ_d are the real roots of an irreducible polynomial over \mathbb{Q} , e.g., $P_d(x) := \prod_{j=1}^d (x - 2j + 1) - 1$. The general principle of this construction has been elaborated in detail by Temlyakov in his book [352, IV.4] based on results on algebraic number theory, see Borevich, Shafarevich [40], Gruber, Lekkerkerker [155], or Skriganov [302]. Let us also refer to [185] for a detailed exposition of the construction principle based on the above mentioned references. The polynomial P_d has a striking disadvantage from a numerical analysis point of view, namely that the real roots of the polynomials grow with d and therefore the entries in A get huge due to the Vandermonde structure. In fact, sticking to the structure (8.9), it seems to be a crucial task to find proper irreducible polynomials with real roots of small modulus. In

[352, IV.4] Temlyakov proposed the use of rescaled Chebyshev polynomials Q_d in dimensions $d = 2^\ell$. To be more precise we use for $x \in [-2, 2]$

$$Q_d(x) = 2T_d(x/2) \quad \text{with} \quad T_d(\cdot) := \cos(d \arccos(\cdot)).$$

The polynomials Q_d belong to $\mathbb{Z}[x]$ and have leading coefficient 1. Its roots are real and given by

$$\xi_k = 2 \cos \left(\frac{\pi(2k-1)}{2d} \right) \quad , \quad k = 1, \dots, d. \quad (8.10)$$

Let us denote the Vandermonde matrix (8.9) with the scaled Chebyshev roots (8.10) by T and call the corresponding lattice $\Gamma_T = T(\mathbb{Z}^d)$ a *Chebyshev lattice*. It turned out recently, see [186], that a Chebyshev lattice is always orthogonal. To be more precise we have the following result.

Theorem 8.16. *The d -dimensional Chebyshev lattice $\Gamma_T = T(\mathbb{Z}^d)$ is orthogonal. In particular, there exists a lattice representation $\tilde{T} = TS$ with $S \in SL_d(\mathbb{Z})$ such that*

$$(i) \quad \tilde{T}_{k,\ell} \in [-2, 2] \text{ for } k, \ell = 1, \dots, d \text{ and}$$

$$(ii) \quad \tilde{T}^* \tilde{T} = \text{diag}(d, 2d, \dots, 2d).$$

However, Chebyshev-polynomials are not always irreducible over \mathbb{Q} . In fact, the polynomials Q_d are irreducible if and only if $d = 2^\ell$ [352, IV.4]. Hence, a Chebyshev lattice Γ_T is admissible in the sense of Theorem 8.15 if and only if $d = 2^\ell$. In that case we call Γ_T a *Chebyshev-Frolov lattice* and obtain the following corollary.

Corollary 8.17. *If $d = 2^\ell$ for some $\ell \in \mathbb{N}$ the Chebyshev-Frolov lattice $\Gamma_T = T(\mathbb{Z}^d)$ and its dual lattice are both orthogonal and admissible (in the sense of Theorem 8.15). In particular, there is a lattice representation for Γ given by $\tilde{T} = QD$ with a diagonal matrix $D = \text{diag}(\sqrt{d}, \sqrt{2d}, \dots, \sqrt{2d})$ and an orthogonal matrix Q . For the dual lattice Γ^\perp we have the representation $\tilde{T}^\perp = QD^{-1}$.*

Let $a > 1$ and A a matrix from Lemma 8.15. We consider the cubature formula

$$\Phi(a, A)(f) := (a^d |\det A|)^{-1} \sum_{\mathbf{m} \in \mathbb{Z}^d} f\left(\frac{(A^{-1})^T \mathbf{m}}{a}\right) \quad (8.11)$$

for f with support in $[0, 1]^d$. Clearly, the number N of points of this cubature formula does not exceed $C(A)a^d |\det A|$. In our framework (8.1) the weights λ_i , $i = 1, \dots, m$, are all equal but do not sum up to one in general. The following figure illustrates the construction of the Frolov points.

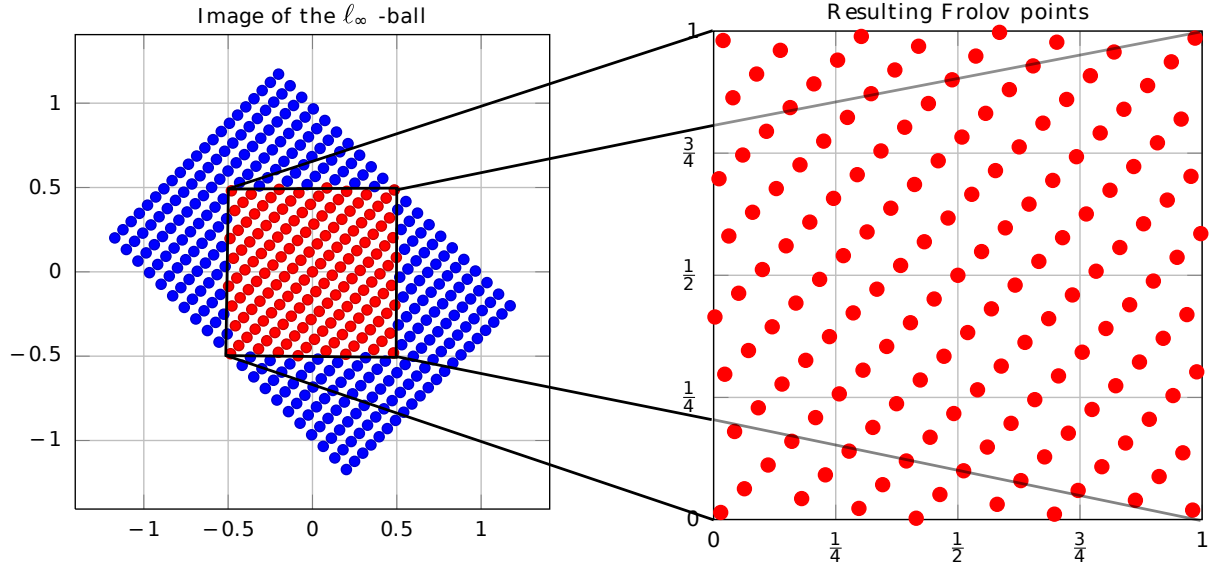


Figure 18: Generating the Frolov points

The first step in application of the Frolov cubature formulas is to apply them to analogs of classes \mathbf{W}_p^r and $\mathbf{B}_{p,\theta}^r$ of respective functions with support in $[0, 1]^d$. Let us first give a precise definition of these classes by using the approach via differences from Lemmas 3.10, 3.11. Let $1 < p < \infty$ and $m > r > 0$ then $\mathbf{W}_p^r(\mathbb{R}^d)$ is the collection of functions $f \in L_1(\mathbb{R}^d)$ such that

$$\|f\|_{\mathbf{W}_p^r(\mathbb{R}^d)}^{(m)} := \left\| \left(\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{r|\mathbf{s}|_1} \mathcal{R}_m^{e(\mathbf{s})}(f, 2^{-\mathbf{s}}, \cdot)^2 \right)^{1/2} \right\|_{L_p(\mathbb{R}^d)}$$

is finite. If $1 \leq p, \theta \leq \infty$ and $m > r > 0$ then $\mathbf{B}_{p,\theta}^r(\mathbb{R}^d)$ is the collection of all functions $f \in L_1(\mathbb{R}^d)$ such that

$$\|f\|_{\mathbf{B}_{p,\theta}^r(\mathbb{R}^d)}^{(m)} := \left[\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{r|\mathbf{s}|_1} \omega_m^{e(\mathbf{s})}(f, 2^{-\mathbf{s}})_{L_p(\mathbb{R}^d)}^\theta \right]^{1/\theta}$$

is finite. Now we consider those functions from $\mathbf{W}_p^r(\mathbb{R}^d)$ which are supported in the cube $[0, 1]^d$, by setting

$$\mathring{\mathbf{W}}_p^r := \{f \in \mathbf{W}_p^r(\mathbb{R}^d) : \text{supp } f \subset [0, 1]^d\}$$

and

$$\mathring{\mathbf{B}}_{p,\theta}^r := \{f \in \mathbf{B}_{p,\theta}^r(\mathbb{R}^d) : \text{supp } f \subset [0, 1]^d\}.$$

For the moment let us use the letter A also for the rescaled Frolov matrix (w.l.g. positive determinant). Then (8.11) can be rewritten to (see [352, Lem. IV.4.6])

$$\frac{1}{\det A} \sum_{\mathbf{m} \in \mathbb{Z}^d} f((A^{-1})^T \mathbf{m}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \mathcal{F}f(A\mathbf{k}), \quad (8.12)$$

which is a consequence of Poisson's summation formula involving the continuous Fourier transform given by

$$\mathcal{F}f(\mathbf{y}) := \int_{-\infty}^{\infty} f(\mathbf{x}) e^{-2\pi i(\mathbf{y}, \mathbf{x})} d\mathbf{x}.$$

Before commenting on this identity let us mention that it immediately gives a formula for the integration error since $\mathcal{F}f(0) = \int_{[-1,1]^d} f(\mathbf{x}) d\mathbf{x}$. In (8.12) the left-hand side is a finite sum due to the support assumption on f . However, the right-hand side does not have to be unconditionally convergent. However, any convergent method of summation of the Fourier series can be used to establish the above identity. For details we refer to [388, Cor. 3.2]. The formula (8.12) is the heart of the matter for the analysis of the method in spaces $\mathring{\mathbf{W}}_p^r$ and $\mathring{\mathbf{B}}_{p,\theta}^r$. The first step consists in proving a counterpart of Theorem 8.14 for spaces $\mathring{\mathbf{W}}_p^r$ and $\mathring{\mathbf{B}}_{p,\theta}^r$, see [352, 365, 388].

Theorem 8.18. (i) *For each $1 < p < \infty$ and $r > \max\{1/p, 1/2\}$ we have*

$$\Phi(a, A)(\mathring{\mathbf{W}}_p^r) \asymp a^{-rd}(\log a)^{(d-1)/2}, \quad a > 1.$$

(ii) *For each $1 \leq p, \theta \leq \infty$ and $r > 1/p$, we have*

$$\Phi(a, A)(\mathring{\mathbf{B}}_{p,\theta}^r) \asymp a^{-rd}(\log a)^{(d-1)(1-1/\theta)}, \quad a > 1.$$

(iii) *For each $1 \leq p < \infty$ we have*

$$\Phi(a, A)(\mathring{\mathbf{B}}_{p,1}^{1/p}) \asymp a^{-d/p}, \quad a > 1.$$

Note, that in case $\theta = 1$ the log term disappears and multivariate cubature shows the same asymptotical behavior as univariate quadrature in this setting.

The case of small smoothness

Here we deal with the cubature of functions from classes \mathbf{W}_p^r where $2 < p < \infty$ and $1/p < r \leq 1/2$. Using Frolov's cubature formula we can prove a multivariate counterparts of the upper bounds in Theorem 8.13. Indeed using the modified Frolov cubature $\Phi(a, A)(f)$ we obtain

Theorem 8.19. *Let $2 < p < \infty$ and $1/p < r \leq 2$. Then the following bounds hold true for $a > 2$*

$$\Phi(a, A)(\mathring{\mathbf{W}}_p^r) \lesssim \begin{cases} a^{-rd}(\log a)^{(d-1)(1-r)}, & 1/p < r < 1/2; \\ a^{-rd}(\log a)^{(d-1)/2} \sqrt{\log \log a}, & r = 1/2. \end{cases}$$

In contrast to the Fibonacci cubature rules the lower bounds in this situation are still open.

8.6 Modifications of Frolov's method

Frolov's method works well for functions with zero boundary condition, see Theorem 8.18 above. In order to treat periodic functions from $\mathbf{B}_{p,\theta}^r$ and \mathbf{W}_p^r with Frolov's method we study the following recently developed modification of the algorithm, which has been introduced in [239]. Let $\psi : \mathbb{R}^d \rightarrow [0, \infty)$ be a compactly supported function ($\text{supp } \psi \subset \Omega$) such that

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \psi(\mathbf{x} + \mathbf{k}) = 1, \quad \mathbf{x} \in \mathbb{R}^d.$$

Then for any 1-periodic function in each component we have

$$\int_{\mathbb{R}^d} \psi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \sum_{\mathbf{k} \in \mathbb{Z}^d} \int_{[0,1]^d} \psi(\mathbf{x} + \mathbf{k}) f(\mathbf{x} + \mathbf{k}) d\mathbf{x} = \int_{[0,1]^d} f(\mathbf{x}) \sum_{\mathbf{k} \in \mathbb{Z}^d} \psi(\mathbf{x} + \mathbf{k}) d\mathbf{x} = \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x}.$$

Let \mathbb{X}_a be the Frolov points generated by the rescaled matrix $\frac{1}{a}(A^{-1})^T$. The modified cubature formula is then

$$\Phi(a, A, \psi)(f) := \Phi(a, A)(\psi f) = (a^d \det A)^{-1} \sum_{\mathbf{x} \in \mathbb{X}_a \cap \Omega} \psi(\mathbf{x}) f(\{\mathbf{x}\}),$$

where $\{\mathbf{x}\} \in [0, 1)^d$ denotes the fractional part of the components in \mathbf{x} . Proving the boundedness of the operator $M_\psi : \mathbf{W}_p^r \rightarrow \mathbf{W}_p^r(\mathbb{R}^d)$ which maps $f \mapsto \psi f$, see [239], one ends up with

$$\Phi(a, A)(\mathring{\mathbf{W}}_p^r) \asymp \Phi(a, A, \psi)(\mathbf{W}_p^r) \quad , \quad a > 1,$$

as well as

$$\Phi(a, A)(\mathring{\mathbf{B}}_{p,\theta}^r) \asymp \Phi(a, A, \psi)(\mathbf{B}_{p,\theta}^r) \quad , \quad a > 1.$$

which implies Theorem 8.14.

Change of variable

Let us also mention a classical modification of the algorithm which goes back to Bykovskii [46], see also [352, 365] for details.

Let ℓ be a natural number. We define the following functions

$$\psi_\ell(u) = \begin{cases} \int_0^u t^\ell (1-t)^\ell dt \bigg/ \int_0^1 t^\ell (1-t)^\ell dt, & u \in [0, 1], \\ 0 & u < 0, \\ 1 & u > 1; \end{cases}$$

For continuous functions of d variables we define the cubature formulas

$$\Phi(a, A, \ell)(f) := \Phi(a, A) \left(f(\psi_\ell(x_1), \dots, \psi_\ell(x_d)) \prod_{i=1}^d \psi'_\ell(x_i) \right). \quad (8.13)$$

Analyzing this method boils down to mapping properties of the operator

$$T_\ell : f \mapsto f(\psi_\ell(x_1), \dots, \psi_\ell(x_d)) \prod_{i=1}^d \psi'_\ell(x_i)$$

within \mathbf{W}_p^r and $\mathbf{B}_{p,\theta}^r$. This has been studied for spaces \mathbf{W}_p^r with $r \in \mathbb{N}$ in [352], [365] and for spaces $\mathbf{B}_{p,\theta}^r$, $r > 1/p$, in [115]. The full range of relevant spaces, including fractional smoothness, is considered in the paper [239]. The final result is

$$\Phi(a, A)(\mathring{\mathbf{W}}_p^r) \asymp \Phi(a, A, \ell)(\mathbf{W}_p^r(\mathbb{R}^d)) \quad , \quad a > 1,$$

as well as

$$\Phi(a, A)(\mathring{\mathbf{B}}_{p,\theta}^r) \asymp \Phi(a, A, \ell)(\mathbf{B}_{p,\theta}^r(\mathbb{R}^d)) \quad , \quad a > 1,$$

if ℓ is large (depending on p and r). This shows that the spaces $\mathbf{W}_p^r(\mathbb{R}^d)$, $\mathbf{B}_{p,\theta}^r(\mathbb{R}^d)$ and $\mathbf{W}_p^r, \mathbf{B}_{p,\theta}^r$ also show the rate of convergence given in Theorem 8.14 with respect to κ_m .

Random Frolov

Further modifications of Frolov's method can be found in the recent papers Krieg, Novak [205] and M. Ullrich [387]. The authors define a Monte Carlo integration method M_m^ω with m nodes based on the Frolov cubature formula and obtain the typical $\min\{1/2, 1 - 1/p\}$ gain in the main rate of convergence. In addition, M. Ullrich [387] observed the surprising phenomenon that the logarithm disappears in the rooted mean square error (variance), i.e.

$$\sup_{f \in \mathbf{W}_p^r} (\mathbb{E}|I(f) - M_m^\omega(f)|^2)^{1/2} \lesssim m^{-(r+1-1/p)}.$$

if $1 < p < \infty$ and $r \geq (1/p - 1/2)_+$.

8.7 Quasi-Monte Carlo cubature

Theorems 8.1, 8.3 provide the lower bounds for numerical integration by cubature rules with m nodes. Theorem 8.14 shows that in a large range of smoothness r there exist cubature rules with m nodes, which provide the upper bounds for the minimal error of numerical integration matching the corresponding lower bounds. This means, that in the sense of asymptotical behavior of $\kappa_m(\mathbf{F})$ the problem is solved for a large range of function classes. In the case of functions of two variables optimal cubature rules are very simple – the Fibonacci cubature rules. They represent a special type of cubature rules, so-called quasi-Monte Carlo rules. In the case $d \geq 3$ the optimal (in the sense of order) cubature rules, considered in the previous Section, are constructive but not as simple as the Fibonacci cubature formulas. In fact, for the Frolov cubature formulae (8.11) all weights λ_i , $i = 1, \dots, m$, in (8.1) are equal but do not sum up to one. This means in particular that constant functions would not be integrated exactly by Frolov's method. Equal weights which sum up to one is the main feature of quasi-Monte Carlo integration. In this Section we will discuss the problem of finding optimal quasi-Monte Carlo rules.

Let us start with the following result.

Theorem 8.20. *For $0 < r < 1$ one has*

$$\kappa_m(\mathbf{H}_\infty^r) \asymp m^{-r}(\log m)^{d-1}.$$

The lower bound in the above theorem follows from Theorem 8.3. The upper bound is provided by the Korobov cubature formula. Let $m \in \mathbb{N}$, $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{Z}^d$. We consider the cubature formulas

$$P_m(f, \mathbf{a}) := m^{-1} \sum_{\mu=1}^m f\left(\left\{\frac{\mu a_1}{m}\right\}, \dots, \left\{\frac{\mu a_d}{m}\right\}\right)$$

which are called the Korobov cubature formulas. Those formulas are of a special type. Similar to the Frolov cubature formulas all weights are equal. However, in contrast to (8.11) we have the additional feature that the equal weights sum up to 1. Consequently, the Korobov cubature formulas compute exactly integrals of constant functions.

The above considered Fibonacci cubature formulas represent a special case of this framework. Namely, if $d = 2$, $m = b_n$, $\mathbf{a} = (1, b_{n-1})$ we have

$$P_m(f, \mathbf{a}) = \Phi_n(f).$$

Note that in the case $d > 2$ the problem of finding concrete cubature formulas of the type $P_m(f, \mathbf{a})$ as good as the Fibonacci cubature formulas in the case $d = 2$ is unsolved. However,

in the sequel we will stick to the property that the weights sum up to 1. In other words we are interested in quasi-Monte Carlo cubature formulas of type

$$I_m(f, X_m) := \frac{1}{m} \sum_{i=1}^m f(x^i)$$

and ask whether there is a sequence of point sets $\{X_m\}_m$ such that $I_m(\mathbf{F}, X_m)$ provides asymptotically optimal error bounds for classes of multivariate functions \mathbf{F} . For more details and further references see also the recent survey by Dick, Kuo and Sloan [80].

It is well-known that numerical integration of functions with mixed smoothness is closely related to the discrepancy problem, see Subsection 8.8 below. In a certain sense (duality) classical discrepancy theory corresponds to numerical integration in function classes \mathbf{W}_p^r with $r = 1$. We refer the reader to the survey paper [365] and the book [382] for a detailed discussion of the connection between numerical integration and discrepancy. In the discrepancy problem the points X_m have to be constructed such that they are distributed as evenly as possible over the unit cube. Explicit constructions of well-distributed point sets in the unit cube have been introduced by Sobol [306] and by Faure [121]. Later Niederreiter [240] introduced the general concept of (t, n, d) -nets. For such point sets, it has been shown that the star discrepancy (the L_∞ -norm of the discrepancy function), a measure of the distribution properties of a point set, behaves well. We do not want to go into the complicated details of the construction of such point sets. Let us rather take a look on a 2-dimensional example. The 2-dimensional van der Corput point set in base 2 is given by

$$\mathcal{H}_n = \left\{ \left(\frac{t_n}{2} + \frac{t_{n-1}}{2^2} + \dots + \frac{t_1}{2^n}, \frac{s_1}{2} + \frac{s_2}{2^2} + \dots + \frac{s_n}{2^n} \right) : t_i \in \{0, 1\}, s_i = 1 - t_i, i = 1, \dots, n \right\}.$$

and represents a digital $(0, n, 2)$ -net. Here, the quality parameter is $t = 0$ since every dyadic interval with volume 2^{-n} contains exactly 1 point.

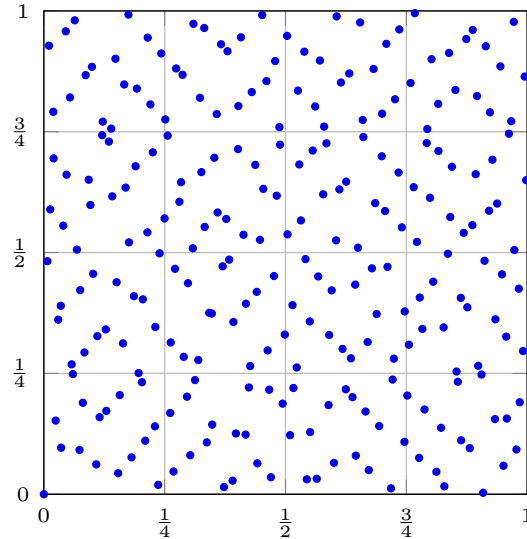


Figure 19: A digital net in $d = 2$ with $N = 256$ points

Based on the work of Hinrichs [173], where Haar coefficients of the discrepancy function have been computed, the following theorem has been shown in [392]

Theorem 8.21. *Let $1 \leq p, \theta \leq \infty$ and $1/p < r < 2$. Then with $m = 2^n$*

$$I_m(\mathbf{B}_{p,\theta}^r(\mathbb{T}^2), \mathcal{H}_n) \asymp m^{-r}(\log m)^{1-1/\theta} \quad , \quad n \in \mathbb{N}.$$

The restriction $r < 2$ comes from the hierarchical decomposition of a 2-variate function into the tensorized Faber-Schauder system. At least the hierarchical decomposition via the Faber-Schauder system, see Subsection 5.5 is problematic for $r > 2$. Note, that the van der Corput quasi-Monte Carlo rule shows the same asymptotic behavior as the Fibonacci cubature formula above in the given range for r , i.e., they are asymptotically optimal. Via embeddings between \mathbf{B} and \mathbf{W} spaces, see Lemma 3.12, we can immediately deduce the corresponding asymptotically optimal result for Sobolev spaces $\mathbf{W}_p^r(\mathbb{T}^2)$ if $1 < p < \infty$ and $2 > r > \max\{1/p, 1/2\}$, see Theorem 8.10(i). What concerns small smoothness we conjecture a counterpart of Proposition 8.13 also in this framework.

Considering the d -dimensional situation, things are much more involved. Based on the b -adic (b large) Chen-Skriganov [55] point construction \mathcal{CS}_n (which we will not discuss in detail) Markhasin [230] showed the following.

Theorem 8.22. *Let $1 \leq p, \theta \leq \infty$ and $1/p < r \leq 1$. Then with $m = b^n$*

$$I_m(\mathbf{B}_{p,\theta}^r, \mathcal{CS}_n) \asymp m^{-r}(\log m)^{(d-1)(1-1/\theta)} \quad , \quad n \in \mathbb{N}.$$

The condition $r \leq 1$ is, of course, not satisfying. In order to prove a d -dimensional counterpart of Theorem 8.21 we need to use higher order digital nets as recently introduced by Dick [77]. Higher order digital nets are designed to achieve higher order convergence rates if the function possesses higher order smoothness, e.g. higher mixed partial derivatives. In fact, classical (t, m, d) nets often admit optimal discrepancy estimates which (roughly) transfer to optimal integration errors within the class \mathbf{W}_p^1 , see Subsection 8.8 below. Several numerical experiments in [175] illustrate this fact. However, for higher order smoothness corresponding statements are not known. What concerns order-2 digital nets \mathcal{DN}_n^2 the following asymptotically optimal results have been shown recently in [175].

Theorem 8.23. *Let $1 \leq p, \theta \leq \infty$ and $1/p < r < 2$. Then with $m = 2^n$*

$$I_m(\mathbf{B}_{p,\theta}^r, \mathcal{DN}_n^2) \asymp m^{-r}(\log m)^{(d-1)(1-1/\theta)} \quad , \quad n \in \mathbb{N}.$$

Compared to the Frolov cubature formulas the restriction $r < 2$ is still unsatisfactory. However, when it comes to the integration of so-called “kink-functions” from mathematical finance like, e.g., integrands of the form $f(t) = \max\{0, t - 1/2\}$ one observes a Hölder-Nikol’skii regularity $r = 2$ if $p = 1$. Hence the above method can take advantage of the maximal regularity of a kink. This is what one observes in numerical experiments as well.

Let us emphasize that the above stated results suffer from the restriction $r < 2$ in case $d \geq 2$. It is not known whether the optimal order of convergence can be achieved by a quasi-Monte Carlo rule in case of higher smoothness. In case $d = 2$ this is the case (Fibonacci). Note, that the modified Frolov method yields the optimal rate for higher smoothness. However, this method is no QMC rule (not even the “pure” Frolov method). Let us mention that there is progress in this direction, see the recent preprints by Goda, Suzuki, Yoshiki [149, 150], where higher order digital nets and corresponding quasi-Monte Carlo rules are used for non-periodic reproducing kernel Hilbert spaces, which extends Theorem 8.23 in case $p = \theta = 2$ towards higher smoothness and non-periodic functions.

8.8 Discrepancy and numerical integration

Classical discrepancy

In this chapter we discussed in detail the problem of numerical integration in the mixed smoothness classes. It is clear that for a formulation of the problem of optimal cubature formulas (optimal numerical integration rules) we need to specify a function class of functions which we numerically integrate. There are many different function classes of interest. In this section we begin with a very simple class, which has a nice geometrical interpretation. The classical L_∞ discrepancy (star-discrepancy) of a set $X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\}$ is defined as follows

$$D(X_m, L_\infty) := \sup_{\mathbf{y} \in [0,1]^d} \left| \frac{1}{m} \sum_{j=1}^m \chi_{[0,\mathbf{y}]}(\mathbf{x}^j) - \int_{[0,1]^d} \chi_{[0,\mathbf{y}]}(\mathbf{x}) d\mathbf{x} \right|.$$

Here, $\chi_{[0,\mathbf{y}]}(\mathbf{x}) := \prod_{j=1}^d \chi_{[0,y_j]}(x_j)$, $y_j \in [0,1]$, $j = 1, \dots, d$, where $\chi_{[0,y]}(x)$, $y \in [0,1]$, is a characteristic function of an interval $[0, y]$. Thus in this case the discrepancy problem is exactly the problem of numerical integration of functions from the class $\chi^d := \{\chi_{[0,\mathbf{y}]}(\cdot)\}_{\mathbf{y} \in [0,1]^d}$ by cubature formulas with equal weights $1/m$.

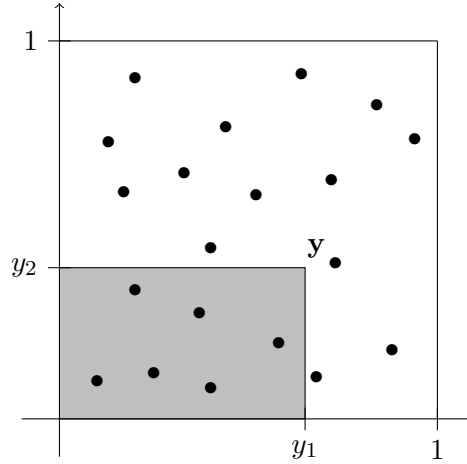


Figure 20: The discrepancy function

The description of the class χ^d is simple – the functions in χ^d are labeled by $\mathbf{y} \in [0,1]^d$. This allows us to consider the L_∞ norm, with respect to \mathbf{y} , of the error which represents a “worst case”. In the case of the L_p norms, $p < \infty$, of the error we may speak of an “average case”. For $1 \leq p < \infty$ we define

$$D(X_m, L_p) := \left\| \frac{1}{m} \sum_{j=1}^m \chi_{[0,\mathbf{y}]}(\mathbf{x}^j) - \int_{[0,1]^d} \chi_{[0,\mathbf{y}]}(\mathbf{x}) d\mathbf{x} \right\|_p.$$

It is easy to check that in this case

$$D(X_m, L_p) = \sup_{f \in \mathbf{W}_{p'}^1} \left| \frac{1}{m} \sum_{j=1}^m f(\mathbf{x}^j) - \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} \right|,$$

where $\dot{\mathbf{W}}_{p'}^1$ consists of the functions $f(\mathbf{x})$ representable in the form

$$f(\mathbf{x}) = \int_{[0,1]^d} \chi_{[\mathbf{0}, \mathbf{y}]}(\mathbf{x}) \varphi(\mathbf{y}) d\mathbf{y}, \quad \|\varphi\|_{p'} \leq 1.$$

Again, in the case $1 \leq p < \infty$ the discrepancy problem is exactly the problem of numerical integration of functions from the class $\dot{\mathbf{W}}_{p'}^1$. Note the duality between p in the discrepancy and p' in the class. Thus, the classical problem of the L_p discrepancy coincides with the problem of numerical integration by the Quasi-Monte Carlo rules of the class $\dot{\mathbf{W}}_{p'}^1$ of smoothness 1. Usually, we consider the whole range of smoothness parameters r in numerical integration. Denote $\Omega_d := [0, 1]^d$. It will be convenient for us to consider the class $\dot{\mathbf{W}}_p^r := \dot{\mathbf{W}}_p^r(\Omega_d)$ consisting of the functions $f(\mathbf{x})$ representable in the form

$$f(\mathbf{x}) = \int_{\Omega_d} B_r(\mathbf{y}, \mathbf{x}) \varphi(\mathbf{y}) d\mathbf{y}, \quad \|\varphi\|_p \leq 1,$$

where

$$B_r(\mathbf{y}, \mathbf{x}) := \prod_{j=1}^d ((r-1)!)^{-1} (y_j - x_j)_+^{r-1}, \quad \mathbf{y}, \mathbf{x} \in \Omega_d. \quad (8.14)$$

Note that in the case $r = 1$ we have $B_1(\mathbf{y}, \mathbf{x}) = \chi_{[\mathbf{0}, \mathbf{y}]}(\mathbf{x})$. In connection with the definition of the class $\dot{\mathbf{W}}_p^r$ we remark here that for the error of the cubature formula $\Lambda_m(\cdot, X_m)$ with weights $\Lambda_m = (\lambda_1, \dots, \lambda_m)$ and nodes $X_m = (\mathbf{x}^1, \dots, \mathbf{x}^m)$ the following relation holds. Let

$$\left| \Lambda_m(f, X_m) - \int_{\Omega_d} f(\mathbf{x}) d\mathbf{x} \right| =: R_m(\Lambda_m, X_m, f),$$

then

$$\begin{aligned} \Lambda_m(\dot{\mathbf{W}}_p^r, X_m) &:= \sup_{f \in \dot{\mathbf{W}}_p^r} R_m(\Lambda_m, X_m, f) \\ &= \left\| \sum_{\mu=1}^m \lambda_\mu B_r(\mathbf{y}, \xi^\mu) - \prod_{j=1}^d (t_j^r / r!) \right\|_{p'} =: D_r(X_m, \Lambda_m, d)_{p'}. \end{aligned} \quad (8.15)$$

The quantity $D_r(X_m, \Lambda_m, d)_q$ in the case $r = 1$, $\Lambda_m = (1/m, \dots, 1/m)$ is the classical L_q discrepancy of the set of points X_m . In the case $\Lambda_m = (1/m, \dots, 1/m)$ we denote $D_r(X_m, d)_q := D_r(X_m, (1/m, \dots, 1/m), d)_q$ and call it the r -discrepancy (see [353] and [365]). Thus, the quantity $D_r(X_m, \Lambda_m, d)_q$ defined in (8.15) is a natural generalization of the concept of discrepancy

$$D(X_m, L_q) = D(X_m, d)_q := D_1(X_m, d)_q. \quad (8.16)$$

This generalization contains two ingredients: general weights Λ_m instead of the special case of equal weights $(1/m, \dots, 1/m)$ and any natural number r instead of $r = 1$. We note that in approximation theory we usually study the whole scale of smoothness classes rather than an individual smoothness class. The above generalization of discrepancy for arbitrary positive integer r allows us to study the question: How does smoothness r affect the rate of decay of generalized discrepancy?

The following result (see [365]) connects the optimal errors of numerical integration for classes $\dot{\mathbf{W}}_p^r$ and \mathbf{W}_p^r . Recall the definition of the quantities κ_m in (8.2) and (8.20).

Theorem 8.24. *Let $1 \leq p \leq \infty$. Then for $r \in \mathbb{N}$*

$$\kappa_m(\dot{\mathbf{W}}_p^r) \asymp \kappa_m(\mathbf{W}_p^r). \quad (8.17)$$

General setting

Let us discuss two problems: (I) numerical integration for a function class defined by a kernel $K(\mathbf{x}, \mathbf{y})$; (II) discrepancy with respect to the collection of functions $\{K(\cdot, \mathbf{y})\}_{\mathbf{y} \in [0,1]^d}$. As for Problem (I) let $1 \leq q \leq \infty$. We define a set \mathcal{K}_q of kernels possessing the following properties: Let $K(\mathbf{x}, \mathbf{y})$ be a measurable function on $\Omega^1 \times \Omega^2$. We assume that for any $\mathbf{x} \in \Omega^1$ we have $K(\mathbf{x}, \cdot) \in L_q(\Omega^2)$; for any $\mathbf{y} \in \Omega^2$ the $K(\cdot, \mathbf{y})$ is integrable over Ω^1 and $\int_{\Omega^1} K(\mathbf{x}, \cdot) d\mathbf{x} \in L_q(\Omega^2)$. For a kernel $K \in \mathcal{K}_{p'}$ we define the class

$$\mathbf{W}_p^K := \left\{ f : f = \int_{\Omega^2} K(\mathbf{x}, \mathbf{y}) \varphi(\mathbf{y}) d\mathbf{y}, \quad \|\varphi\|_{L_p(\Omega^2)} \leq 1 \right\}. \quad (8.18)$$

Then each $f \in \mathbf{W}_p^K$ is integrable on Ω^1 (by Fubini's theorem) and defined at each point of Ω^1 . We denote for convenience

$$J(\mathbf{y}) := J_K(\mathbf{y}) := \int_{\Omega^1} K(\mathbf{x}, \mathbf{y}) d\mathbf{x}.$$

For a cubature formula $\Lambda_m(\cdot, X_m)$ we have

$$\begin{aligned} \Lambda_m(\mathbf{W}_p^K, X_m) &= \sup_{\|\varphi\|_{L_p(\Omega^2)} \leq 1} \left| \int_{\Omega^2} (J(\mathbf{y}) - \sum_{\mu=1}^m \lambda_\mu K(\mathbf{x}^\mu, \mathbf{y})) \varphi(\mathbf{y}) d\mathbf{y} \right| \\ &= \left\| J(\cdot) - \sum_{\mu=1}^m \lambda_\mu K(\mathbf{x}^\mu, \cdot) \right\|_{L_{p'}(\Omega^2)}. \end{aligned} \quad (8.19)$$

We use the a similar definition as above of the error of optimal cubature formula with m nodes for a class \mathbf{W}

$$\kappa_m(\mathbf{W}) := \inf_{\lambda_1, \dots, \lambda_m; \mathbf{x}^1, \dots, \mathbf{x}^m} \Lambda_m(\mathbf{W}, X_m). \quad (8.20)$$

Thus, by (8.19)

$$\kappa_m(\mathbf{W}_p^K) = \inf_{\lambda_1, \dots, \lambda_m; \mathbf{x}^1, \dots, \mathbf{x}^m} \left\| J(\cdot) - \sum_{\mu=1}^m \lambda_\mu K(\mathbf{x}^\mu, \cdot) \right\|_{L_{p'}(\Omega^2)}. \quad (8.21)$$

Note, that the error of numerical integration of those classes is closely related to the m -term approximation of a special function $J(\cdot)$ with respect to a dictionary $\{K(\mathbf{x}, \cdot)\}_{\mathbf{x} \in \Omega^1}$. This leads to interesting applications of the recently developed theory of greedy approximation (cf. Section 7 above) in numerical integration. We refer the reader to [365, Sect. 2].

Let us now consider Problem (II). The following definition is from [373].

Definition 8.25. *The (K, q) -discrepancy of a cubature formula Λ_m with nodes $X_m = \{\mathbf{x}^1, \dots, \mathbf{x}^m\}$ and weights $\Lambda_m = (\lambda_1, \dots, \lambda_m)$ is defined as*

$$D(\Lambda_m, K, q) := \left\| \int_{[0,1]^d} K(\mathbf{x}, \mathbf{y}) d\mathbf{x} - \sum_{\mu=1}^m \lambda_\mu K(\mathbf{x}^\mu, \mathbf{y}) \right\|_{L_q([0,1]^d)}.$$

The particular case $K(\mathbf{x}, \mathbf{y}) := \chi_{[0,\mathbf{y}]}(\mathbf{x}) := \prod_{j=1}^d \chi_{[0,y_j]}(x_j)$ leads to the classical concept of the L_q -discrepancy. Therefore, we get the following “duality” between numerical integration and discrepancy (in a very general form)

$$D(X_m, \Lambda_m, K, p') = \Lambda_m(\mathbf{W}_p^K, X_m).$$

Certainly, in the above definitions we can replace the L_p space by more general Banach spaces X . Then we still have “duality” between discrepancy in the norm of X and numerical integration of classes defined in X' .

Let us define the minimal weighted r -discrepancy by

$$D_r^w(m, d)_q := \inf_{X_m, \Lambda_m} D_r(X_m, \Lambda_m, d)_q \quad , \quad 1 \leq q \leq \infty ,$$

whereas the minimal r -discrepancy is given by

$$D_r(m, d)_q := \inf_{X_m} D_r(X_m, d)_q \quad , \quad 1 \leq q \leq \infty ,$$

with the minimal (classical) discrepancy $D(m, d)_q := D_1(m, d)_q$ as a special case. Clearly, it holds

$$D_r^w(m, d)_q \leq D_r(m, d)_q .$$

Historical remarks on discrepancy

To begin with let us mention that there exist several monographs and surveys on discrepancy theory, see [27, 113, 232] to mention just a few. We would like to point out the following milestones regarding upper and lower bounds for the minimal discrepancy.

Lower bounds. Let us start with K. Roth [280] who proved in 1954 that

$$D(m, d)_2 \geq C(d)m^{-1}(\log m)^{(d-1)/2} . \quad (8.22)$$

In 1972 W. Schmidt [289] proved

$$D(m, 2)_\infty \geq Cm^{-1} \log m . \quad (8.23)$$

In 1977 W. Schmidt [290] proved

$$D(m, d)_q \geq C(d, q)m^{-1}(\log m)^{(d-1)/2}, \quad 1 < q \leq \infty . \quad (8.24)$$

In 1981 G. Halász [159] proved

$$D(m, d)_1 \geq C(d)m^{-1}(\log m)^{1/2} . \quad (8.25)$$

The following conjecture has been formulated in [27] as an excruciatingly difficult great open problem.

Conjecture 8.26. *We have for $d \geq 3$*

$$D(m, d)_\infty \geq C(d)m^{-1}(\log m)^{d-1} .$$

This problem is still open. Recently, in 2008, D. Bilyk and M. Lacey [36] and D. Bilyk, M. Lacey, and A. Vagharshakyan [37] proved

$$D(m, d)_\infty \geq C(d)m^{-1}(\log m)^{(d-1)/2+\delta(d)} \quad (8.26)$$

with some $0 < \delta(d) < 1/2$, which essentially improved on the lower bound

$$D(m, d)_\infty \geq C(d)m^{-1}(\log m)^{(d-1)/2} \left(\frac{\log \log m}{\log \log \log m} \right)^{\frac{1}{2d-2}} ,$$

see Baker [12]. The approach in [36, 37] is based on an improved version of the Small Ball Inequality, see Subsection 2.6 above. The conjectured inequality (2.29) would imply the stronger lower bound $C(d)m^{-1}(\log m)^{d/2}$. Compared to Conjecture 8.26 there would still be a large gap in the power of the logarithm if $d > 2$. Note also the connection to metric entropy in L_∞ , see Conjecture 6.16, and the Small Ball Problem in probability theory, see Subsection 6.4.

There seems to be some further progress in connection with Conjecture 8.26. M. Levin [215, 216, 217] recently proved that several widely used point constructions (like digital nets, Halton points, Frolov lattice) satisfy the lower bound proposed in Conjecture 8.26. The method of proof is deeply involved and uses nontrivial tools from algebraic number theory.

The first result in estimating the weighted r -discrepancy was obtained in 1985 by V.A. Bykovskii [46]

$$D_r^w(m, d)_2 \geq C(r, d)m^{-r}(\log m)^{(d-1)/2}. \quad (8.27)$$

This result is a generalization of Roth's result (8.22). The generalization of Schmidt's result (8.24) was obtained by V.N. Temlyakov [345] in 1990

$$D_r^w(m, d)_q \geq C(r, d, q)m^{-r}(\log m)^{(d-1)/2}, \quad 1 < q \leq \infty. \quad (8.28)$$

In 1994 V.N. Temlyakov [353] proved that for r even integers we have for the minimal r -discrepancy

$$D_r(m, d)_\infty \geq C(r, d)m^{-r}(\log m)^{d-1}. \quad (8.29)$$

This result encourages us to formulate the following generalization of Conjecture 8.26.

Conjecture 8.27. *For all $d, r \in \mathbb{N}$ we have*

$$D_r^w(m, d)_\infty \geq C(r, d)m^{-r}(\log m)^{d-1}.$$

The above lower estimates for $D_1^w(m, d)_q$ are formally stronger than the corresponding estimates for $D(m, d)_q$ because in $D_1^w(m, d)_q$ we are in addition optimizing over the weights Λ_m .

Upper bounds. We now present the upper estimates for the discrepancy in various settings. In 1956 H. Davenport [67] proved that

$$D(m, 2)_2 \leq Cm^{-1}(\log m)^{1/2}.$$

Other proofs of this estimate were later given by I.V. Vilenkin [394], J.H. Halton and S.K. Zaremba [162], and K. Roth [281]. In 1979 K. Roth [282] proved

$$D(m, 3)_2 \leq Cm^{-1} \log m$$

and in 1980 K. Roth [283] and K.K. Frolov [130] proved

$$D(m, d)_2 \leq C(d)m^{-1}(\log m)^{(d-1)/2}.$$

In 1980 W. Chen [54] (and later in 1994 M. Skriganov [302]) proved

$$D(m, d)_q \leq C(d)m^{-1}(\log m)^{(d-1)/2}, \quad 1 < q < \infty.$$

Upper bounds for the star-discrepancy have been known since 1960. J.M Hammersley [163] and J.H. Halton [161] were the first who gave explicit point constructions proving that

$$D(m, d)_\infty \leq C(d)m^{-1}(\log m)^{d-1}. \quad (8.30)$$

As we will point out below in detail, see Subsection 10.2, a bound like (8.30) is non-trivial only for $m > e^{d-1}$ (without additional knowledge on the involved constant). For many applications this threshold is prohibitively large. What concerns the “preasymptotic range” for m let us mention a surprising (non-constructive) result by S. Heinrich, E. Novak, G. Wasilkowski, H. Woźniakowski [169] from 2001. Via probabilistic arguments based on deep results from the theory of empirical processes (Dudley 1984, Talagrand 1994) and combinatorics (Haussler 1995) the authors in [169] proved the existence of a universal (but unknown) constant $c > 0$ such that for arbitrary dimension $d \in \mathbb{N}$

$$D(m, d)_\infty \leq c\sqrt{d/m} \quad , \quad m \in \mathbb{N}. \quad (8.31)$$

Essentially, it is shown that m points drawn uniformly at random in $[0, 1]^d$ satisfy the upper bound in (8.31) with non-zero probability. A simple proof of the above result is given by Aistleitner [1] with constant $c = 10$. Note, that the expected star-discrepancy of a random point set is of order $\sqrt{d/m}$ as shown by Doerr [111]. Hinrichs [172] proved a lower bound for $D(m, d)_\infty$ which is also polynomial in d/m . In fact, his proof shows that this lower bound is also valid for $D_1^w(m, d)_\infty$.

Let us also mention the following constructive result. Greedy approximation techniques allows us to build constructive sets X_m and $X_m(p)$, $1 \leq p < \infty$, such that for $d, m \geq 2$ (see [368], pp. 402–403)

$$\begin{aligned} D(X_m, L_\infty) &\leq c_1 d^{3/2} (\max\{\ln d, \ln m\})^{1/2} m^{-1/2} \\ D(X_m(p), L_p) &\leq c_2 p^{1/2} m^{-1/2}, \quad 1 \leq p < \infty, \end{aligned} \quad (8.32)$$

with effective absolute constants c_1 and c_2 .

We finally comment on upper bounds for the weighted r -discrepancy. The estimate in Theorem 8.14 above together with Theorem 8.24 implies

$$D_r^w(m, d)_\infty \leq C(r, d) m^{-r} (\log m)^{d-1}, \quad r \geq 2.$$

It is clear that upper estimates for $D(m, d)_q$ are stronger than the same upper estimates for $D_1^w(m, d)_q$. Let us finally emphasize that the Smolyak nets are very poor from the point of view of discrepancy as it was shown by N. Nauryzbayev and N. Temirgaliyev (see [234] and [235]).

8.9 Open problems and historical comments

Let us begin with a list of open problems which we will partly discuss below.

Open problem 8.1 Find the right order of the optimal error of numerical integration $\kappa_m(\mathbf{W}_1^r)$ if $r \geq 1$ (see Conjecture 8.28 below).

Open problem 8.2 Find the right order of the optimal error of numerical integration $\kappa_m(\mathbf{W}_p^r)$ in the range of small smoothness (see Conjecture 8.30).

Open problem 8.3 Find the right order of the optimal error of numerical integration $\kappa_m(\mathbf{W}_\infty^r)$ if $r > 0$ (see Conjecture 8.29 below).

Open problem 8.4 Find the right order of the star-discrepancy, see (8.30) and Conjecture 8.26 above.

Let us give some historical comments on the subject which go back to the 1950s. We begin with the lower estimates for cubature formulas. The results from Theorem 8.1 have forerunners. For the class \mathbf{W}_2^r Theorem 8.1 was established in [46] by a different method. Theorem 8.1 was proved in [345]. In the case $\theta = \infty$ the lower bound in Theorem 8.3 has been observed by Bakhvalov [18] in 1972. Theorem 8.7 was proved in [353]. For recent new proofs of Theorems 8.1, 8.3 we refer to [388] and [109]. Concerning lower bounds there are several open problems. Let us formulate them as conjectures, see [365].

Conjecture 8.28. *For any $d \geq 2$ and any $r > 1$ we have*

$$\kappa_m(\mathbf{W}_1^r) \geq C(r, d)m^{-r}(\log m)^{d-1}.$$

Conjecture 8.29. *For any $d \geq 2$ and any $r > 0$ we have*

$$\kappa_m(\mathbf{W}_\infty^r) \geq C(r, d)m^{-r}(\log m)^{(d-1)/2}.$$

Also important is the missing lower bounds for the case of small smoothness in the Sobolev case. There is so far only a result for the special case of the Fibonacci cubature formula, see Proposition 8.13 and [347] which supports the following conjecture.

Conjecture 8.30. *Let $d \geq 2$, $2 < p < \infty$ and $1/p < r \leq 1/2$. Then we have*

$$\kappa_m(\mathbf{W}_p^r) \gtrsim \begin{cases} m^{-r}(\log m)^{(d-1)(1-r)}, & 1/p < r < 1/2; \\ m^{-r}(\log m)^{(d-1)/2}\sqrt{\log \log m}, & r = 1/2. \end{cases}$$

A first step could be to establish such a lower bound for the “pure” Frolov method with respect to the class of Sobolev functions \mathbf{W}_p^r supported in the cube $[0, 1]^d$ which is of course smaller than \mathbf{W}_p^r .

We turn to the upper bounds. The first result in this direction was obtained by N.M. Korobov [202] in 1959. He used the cubature formulas $P_m(f, \mathbf{a})$. His results lead to the following bound

$$\kappa_m(\mathbf{W}_1^r) \leq C(r, d)m^{-r}(\log m)^{rd}, \quad r > 1. \quad (8.33)$$

In 1959 N.S. Bakhvalov [16] improved (8.33) to

$$\kappa_m(\mathbf{W}_1^r) \leq C(r, d)m^{-r}(\log m)^{r(d-1)}, \quad r > 1.$$

It is worth mentioning that Korobov and Bakhvalov worked with the space \mathbf{E}_d^r . The above mentioned result is a consequence of the embedding $\mathbf{W}_1^r \subset \mathbf{E}_d^r$, see Lemma 3.16 above.

There is vast literature on cubature formulas based on function values at the nodes of number-theoretical nets. We do not discuss this literature in detail because these results do not provide the optimal rate of errors for numerical integration. A typical bound differs by an extra $(\log m)^a$ factor. The reader can find many results of this type in the books [203], [204], [179], [352]. For the case of small smoothness see [332]. An interesting method of building good Korobov’s cubature formulas was suggested by S. M. Voronin and N. Temirgaliev in [395]. It is based on the theory of divisors. This method was further developed in a number of papers by N. Temirgaliev and his students [317]–[318], [408], [301], [11].

The first best possible upper estimate for the classes \mathbf{W}_p^r was obtained by N.S. Bakhvalov [17] in 1963. He proved in the case $d = 2$ that

$$\kappa_m(\mathbf{W}_2^r) \leq C(r)m^{-r}(\log m)^{1/2}, \quad r \in \mathbb{N}. \quad (8.34)$$

N.S. Bakhvalov used the Fibonacci cubature formulas. Propositions 8.11(i), 8.13 and 8.12(i),(ii) have been proved in [347] and [353]. Proposition 8.11(ii) in the particular case $\theta = \infty$, $1 \leq p \leq \infty$ was proved in [347] (see also [352], Ch. 4, Theorem 2.6). The same method gives the statement of Theorem 8.10(ii) for all remaining cases including Proposition 8.12(iii), see [109].

A. Hinrichs and J. Oettershagen [177] showed that the Fibonacci points are the globally optimal point set of size N for the quasi-Monte Carlo integration in $\mathbf{W}_2^r(\mathbb{T}^2)$ for some numbers $N \in \mathbb{N}$.

In 1976 K.K. Frolov [128] used the Frolov cubature formulas to extend (8.34) to the case $d > 2$:

$$\kappa_m(\mathring{\mathbf{W}}_2^r) \leq C(r, d)m^{-r}(\log m)^{(d-1)/2}, \quad r \in \mathbb{N}.$$

In 1985 V.A. Bykovskii [46] proved the equivalence

$$\kappa_m(\mathring{\mathbf{W}}_p^r) \asymp \kappa_m(\mathbf{W}_p^r) \tag{8.35}$$

in the case $p = 2$ and got the upper bound in Theorem 8.14(i) in the case $p = 2$. Bykovskii also used the Frolov cubature formulas. Relation (8.35) for $1 < p < \infty$ and $r \in \mathbb{N}$, and its extension to classes \mathbf{H}_p^r , $1 \leq p \leq \infty$, $r > 1/p$ have been proved in [345] and [353] (see also [365], [352, Chapt. IV.4]), and in [115]. Note, that the matter reduces in proving the boundedness of a certain change of variable operator between the respective spaces. For a complete solution of this problem we refer to [239]. Theorem 8.20 is taken from [352]. The upper bounds in Theorems 8.14(ii) and 8.18(ii) were proved by V.V. Dubinin [114, 115] in 1992 and 1997. The upper bound in Theorem 8.19 has been proved recently by M. Ullrich and T. Ullrich [388], see also [239].

The Frolov cubature formulas [129] (see also [114, 115, 302] and the recent papers [388, 239]) give the following bound

$$\kappa_m(\mathbf{W}_1^r) \leq C(r, d)m^{-r}(\log m)^{d-1}, \quad r > 1.$$

Thus the lower estimate in Conjecture 8.28 is the best possible.

In 1994 M.M. Skriganov [302] proved the following estimate

$$\kappa_m(\mathring{\mathbf{W}}_p^r) \leq C(r, d, p)m^{-r}(\log m)^{(d-1)/2}, \quad 1 < p \leq \infty, \quad r \in \mathbb{N}.$$

This estimate combined with (8.35) implies

$$\kappa_m(\mathbf{W}_p^r) \lesssim m^{-r}(\log m)^{(d-1)/2}, \quad 1 < p \leq \infty, \quad r \in \mathbb{N}.$$

Theorems 8.18 (for $\theta < \infty$), 8.19 have been proved recently in [388], see also [239] for the extension to periodic and non-periodic spaces on the cube $[0, 1]^d$. Together with Theorems 8.1, 8.3 they imply Theorem 8.14.

The lower bounds in Theorem 8.8 have been proved in [109], see also [388]. These results can be also proven with a different technique, see the recent paper [371]. This technique even shows that the lower bound will not get smaller when adding $2^{\ell-1}$ arbitrary points.

Concerning upper bounds for Smolyak cubature, see also Remark 8.9, we refer to Triebel [382]. The reader can find some further results on numerical integration by Smolyak type methods in papers [322, 319] and [20, 249, 143, 250, 176].

9 Related problems

9.1 Why classes with mixed smoothness?

In this section we briefly discuss development of the hyperbolic cross approximation theory with emphasis put on the development of methods and connections to other areas of research. The theory of the hyperbolic cross approximation began in the papers by Babenko [7] and Korobov [202]. Babenko studied approximation of functions from classes \mathbf{W}_∞^r in L_∞ by the hyperbolic cross polynomials and the Bernstein type inequality for the hyperbolic cross polynomials in the L_∞ norm. Korobov studied numerical integration of functions from classes

$$\mathbf{E}_d^r(C) := \{f \in L_1(\mathbb{T}^d) : |\hat{f}(\mathbf{k})| \leq C \prod_{j=1}^d \max\{1, |k_j|\}^{-r}\},$$

see also Section 3.3 above. One of the Korobov's motivations for studying classes $\mathbf{E}_d^r(C)$ was related to numerical solutions of integral equations. Let $K(x, y)$ be the kernel of the integral operator J_K . Then the kernel of the $(J_K)^d$ is given by

$$K^d(x, y) = \int_{\mathbb{T}^{d-1}} K(x, x_1) K(x_1, x_2) \cdots K(x_{d-1}, y) dx_1 \cdots dx_{d-1}.$$

Smoothness properties of $K(x, y)$ are naturally transformed into mixed smoothness properties of $K(x, x_1) K(x_1, x_2) \cdots K(x_{d-1}, y)$. In the simplest case of $f_j(t)$ satisfying $\|f_j'\|_\infty \leq 1$, $j = 1, 2, \dots, d$ we obtain $\|(f_1(x_1) \cdots f_d(x_d))^{(1,1,\dots,1)}\|_\infty \leq 1$.

It is an *a priori* argument about importance of classes with mixed smoothness. There are other *a priori* arguments in support of importance of classes with mixed smoothness. For instance, recent results on solutions of the Schrödinger equation, which we discussed in the Introduction, give such an argument. There are also strong *a posteriori* arguments in favor of thorough study of classes of mixed smoothness and the hyperbolic cross approximation. These arguments can be formulated in the following general way. Methods developed for the hyperbolic cross approximation are very good in different sense. We discuss this important point in detail, beginning with numerical integration.

It was immediately understood that the trivial generalization of the univariate quadrature formulas with equidistant nodes to cubature formulas with rectangular grids does not work for classes with mixed smoothness. As a result different fundamental methods of numerical integration were constructed: the Korobov cubature formulas (in particular, the Fibonacci cubature formulas), the Smolyak cubature formulas, the Frolov cubature formulas (see Section 8). These nontrivial constructions are very useful in practical numerical integration, especially, when the dimension of the model is moderate (≤ 40). In subsection 9.2 we discuss theoretic results on universality of these methods, which explain such a great success of these methods in applications.

From the general point of view the problem of numerical integration can be seen as a problem of discretization: approximate a “continuous object” – an integral with respect to the Lebesgue measure – by a “discrete object” – a cubature formula. It is a fundamental problem of the discrepancy theory. It is now well understood that the numerical integration of functions with mixed smoothness is closely related to the discrepancy theory (see, for instance, [365]).

Other example of a fundamental problem, which falls into a category of discretization problems is the entropy problem (see Section 6). It turns out that the entropy problem for classes with bounded mixed derivative is equivalent to an outstanding problem of probability theory – the small ball problem. Both of the mentioned above discretization problems are deep and

difficult problems. Some fundamental problems of those areas are still open. The problems that were resolved required new interesting technique. For instance, the Korobov cubature formulas and the Frolov cubature formulas are based on number theoretical constructions. Study of the entropy numbers of classes of functions with mixed smoothness uses deep results from the theory of finite dimensional Banach spaces and geometry, including volume estimates of special convex bodies.

One more fundamental problem of the discretization type is the sampling problem discussed in Section 5. Alike the entropy problem and the problem of numerical integration the sampling problem for classes with mixed smoothness required new techniques. Study of the sampling problem is based on a combination of classical results from harmonic analysis and new embedding type inequalities. Some outstanding problems are still open.

We now briefly comment on some steps in the development of the hyperbolic cross approximation. First sharp in the sense of order results on the behavior of asymptotic characteristics of classes with mixed smoothness were obtained in the L_2 norm. The Hilbertian structure of L_2 , in particular the Parseval identity, was used in those results. In a step from L_2 to L_p , $1 < p < \infty$, different kind of harmonic analysis technique was used. The classical Littlewood-Paley theorem and the Marcinkiewicz multipliers theorem are the standard tools. Later, the embedding type inequalities between the L_q norm of a function and the L_p norms of its dyadic blocs were proved and widely used. At the early stage of the hyperbolic cross approximation the classes \mathbf{W}_p^r and \mathbf{H}_p^r were studied. Many new and interesting phenomena (compared to the univariate approximation) were discovered. Even in the case $1 < p, q < \infty$ the study of asymptotic characteristics of classes \mathbf{W}_p^r and \mathbf{H}_p^r in L_q required new approaches and new techniques. Some of the problems are still open. A step from the case $1 < p, q < \infty$ to the case, when one (or two) of the parameters p, q take extreme values 1 or ∞ turns out to be very difficult. The majority of the problems are still open in this case. Some surprising phenomena were observed in that study. For instance, in many cases the step from a problem for two variables to the problem for $d \geq 3$ variables is not resolved. A number of interesting effects, which required special approaches for being established, was discovered. We only mention a few from a list of new methods, which were used in those discoveries: properties of special polynomials, for example,

$$\left| \sum_{\mathbf{k} \in \Gamma(N), \mathbf{k} > 0} (k_1 \cdots k_d)^{-1} \sin k_1 x_1 \cdots \sin k_d x_d \right| \leq C(d)$$

in a combination with the Nikol'skii duality theorem; the Riesz products; the Small Ball Inequality; the volume estimates of sets of Fourier coefficients of bounded trigonometric polynomials.

The sampling problem, including the numerical integration problem, turns out to be a difficult problem for classes with mixed smoothness. Many outstanding problems in this area are still open. In particular, the problem of numerical integration of classes \mathbf{W}_1^r and \mathbf{W}_∞^r is not resolved yet.

9.2 Universality

In this subsection we illustrate the following general observation. Methods of approximation, which are optimal in the sense of order for the classes with mixed smoothness, are universal for the collection of anisotropic smoothness classes. This gives *a-posteriori* justification for thorough study of classes of functions with mixed smoothness. The phenomenon of saturation is well known in approximation theory [73], Ch.11. The classical example of a saturation method is the Fejér operator for approximation of univariate periodic functions. In the case of the sequence of the Fejér operators K_n , saturation means that the approximation order by

operators K_n does not improve over the rate $1/n$ even if we increase smoothness of functions under approximation. Methods (algorithms) that do not have the saturation property are called unsaturated. The reader can find a detailed discussion of unsaturated algorithms in approximation theory and in numerical analysis in a survey paper [8]. We point out that the concept of smoothness becomes more complicated in the multivariate case than it is in the univariate case. In the multivariate case a function may have different smoothness properties in different coordinate directions. In other words, functions may belong to different anisotropic smoothness classes (see Hölder-Nikol'skii classes below). It is known ([352]) that approximation characteristics of anisotropic smoothness classes depend on the average smoothness and optimal approximation methods depend on anisotropy of classes. This motivated a study in [328] of existence of an approximation method that is good for all anisotropic smoothness classes. This is a problem of existence of a universal method of approximation. We note that the universality concept in learning theory is very important and it is close to the concepts of adaptation and distribution-free estimation in non-parametric statistics ([158], [38], [367]).

We present in this section a discussion of known results on universal cubature formulas. We define the multivariate periodic Hölder-Nikol'skii classes $\mathbf{NH}_p^{\mathbf{r}}$ in the following way. The class $\mathbf{NH}_p^{\mathbf{r}}$, $\mathbf{r} = (r_1, \dots, r_d)$ and $1 \leq p \leq \infty$, is the set of periodic functions $f \in L_p(\mathbb{T}^d)$ such that for each $l_j = [r_j] + 1$, $j = 1, \dots, d$, the following relations hold

$$\|f\|_p \leq 1, \quad \|\Delta_t^{l_j, j} f\|_p \leq |t|^{r_j}, \quad j = 1, \dots, d,$$

where $\Delta_t^{l_j, j}$ is the l_j -th difference with step t in the variable x_j . In the case $d = 1$, $\mathbf{NH}_p^{\mathbf{r}}$ coincides with the standard Hölder class H_p^r .

Let a vector $\mathbf{r} = (r_1, \dots, r_d)$, $r_j > 0$ and a number m be given. Denote $g(\mathbf{r}) := (\sum_{j=1}^d r_j^{-1})^{-1}$. We define numbers $N_j := \max([m^{r_j}], 1)$, $\varrho_j := g(\mathbf{r})/r_j$, $j = 1, \dots, d$ and the cubature formula

$$q_m(f, \mathbf{r}) := q_{\mathbf{N}}(f), \quad \mathbf{N} := (N_1, \dots, N_d).$$

$$q_{\mathbf{N}}(f) := \left(\prod_{j=1}^d N_j \right)^{-1} \sum_{j_d=1}^{N_d} \cdots \sum_{j_1=1}^{N_1} f(2\pi j_1/N_1, \dots, 2\pi j_d/N_d).$$

It is known ([16], [352]) that for $g(\mathbf{r}) > 1/p$

$$\kappa_m(\mathbf{NH}_p^{\mathbf{r}}) \asymp q_m(\mathbf{NH}_p^{\mathbf{r}}, \mathbf{r}) \asymp m^{-g(\mathbf{r})}, \quad 1 \leq p \leq \infty,$$

where

$$q_m(\mathbf{W}, \mathbf{r}) := \sup_{f \in \mathbf{W}} \left| q_m(f, \mathbf{r}) - (2\pi)^{-d} \int_{\mathbb{T}^d} f(\mathbf{x}) d\mathbf{x} \right|.$$

We note that the cubature formula $q_m(\cdot, \mathbf{r})$ depends essentially on the anisotropic class defined by the vector \mathbf{r} . It is known (see [346], [349]) that the Fibonacci cubature formulas (see Section 8 for the definition) are optimal (in the sense of order) among all cubature formulas: for $g(\mathbf{r}) > 1/p$

$$\delta_{b_n}(\mathbf{NH}_p^{\mathbf{r}}) \asymp \Phi_n(\mathbf{NH}_p^{\mathbf{r}}) \asymp b_n^{-g(\mathbf{r})}.$$

Thus, the Fibonacci cubature formulas are universal for the collection $\{\mathbf{NH}_p^{\mathbf{r}} : 1 \leq p \leq \infty, g(\mathbf{r}) > 1/p\}$ in the following sense. The $\Phi_n(\cdot)$ does not depend on the vector \mathbf{r} and the parameter p and provides optimal (in the sense of order) error bound for each class $\mathbf{NH}_p^{\mathbf{r}}$ from the collection.

In the case $d > 2$, just as for $d = 2$, there exist universal cubature formulas for the anisotropic Hölder-Nikol'skii classes (see [346], [352]). At the same time we emphasize that the universal

cubature formulas that we constructed for $d > 2$ have an essentially different character from the Fibonacci cubature formulas. The first cubature formulas of this type have been constructed by K.K. Frolov, see Section 8. Let $Q_n^\ell(f)$ be the cubature formula defined in (8.13). The following result has been obtained in [346] (see also [352]).

Theorem 9.1. *Let $1 < p \leq \infty$, $g(\mathbf{r}) > 1$. Then for all \mathbf{r} such that $r_j \leq (\ell - 1)(1 - 1/p) - 1$, $j = 1, \dots, d$ the following relation holds,*

$$Q_n^\ell(\mathbf{NH}_p^\mathbf{r}) \asymp n^{-g(\mathbf{r})}.$$

This theorem establishes that the cubature formulas Q_n^ℓ are universal for the collection

$$\left\{ \mathbf{NH}_p^\mathbf{r} : 1 < p \leq \infty, g(\mathbf{r}) > 1, r_j \leq (\ell - 1)(1 - 1/p) - 1, j = 1, \dots, d \right\}.$$

We now briefly discuss universality results in the nonlinear approximation. The following observation motivates our interest in the universal dictionary setting. In practice we often do not know the exact smoothness class F where our input function (signal, image) comes from. Instead, we often know that our function comes from a class of certain structure, for instance, anisotropic Hölder-Nikol'skii class. This is exactly the situation we are dealing with in the following universal dictionary setting. We formulate an optimization problem in a Banach space X for a pair of a function class F and a collection \mathbb{D} of bases (dictionaries) \mathcal{D} . We use the following notation from Section 7

$$\begin{aligned} \sigma_m(f, \mathcal{D})_X &:= \inf_{\substack{g_i \in \mathcal{D}, c_i \\ i=1, \dots, m}} \|f - \sum_{i=1}^m c_i g_i\|_X; \\ \sigma_m(F, \mathcal{D})_X &:= \sup_{f \in F} \sigma_m(f, \mathcal{D})_X; \\ \sigma_m(F, \mathbb{D})_X &:= \inf_{\mathcal{D} \in \mathbb{D}} \sigma_m(F, \mathcal{D})_X. \end{aligned}$$

The universality problem is a problem of finding a method of approximation that is good for each class from a given collection of classes. For example, we introduced in [363] the following definition of universal dictionary. Let two collections \mathcal{F} of function classes and \mathbb{D} of dictionaries be given. We say that $\mathcal{D} \in \mathbb{D}$ is universal for the pair $(\mathcal{F}, \mathbb{D})$ if there exists a constant C which may depend only on \mathcal{F} , \mathbb{D} , and X such that for any $F \in \mathcal{F}$ we have

$$\sigma_m(F, \mathcal{D})_X \leq C \sigma_m(F, \mathbb{D})_X.$$

So, if for a collection \mathcal{F} there exists a universal dictionary $\mathcal{D}_u \in \mathbb{D}$, it is an ideal situation. We can use this universal dictionary \mathcal{D}_u in all cases and we know that it adjusts automatically to the best smoothness class $F \in \mathcal{F}$ which contains a function under approximation. Next, if a pair $(\mathcal{F}, \mathbb{D})$ does not allow a universal dictionary we have a trade-off between universality and accuracy.

It was proved in [363] that

$$\sigma_m(\mathbf{NH}_p^\mathbf{r}, \mathbb{O})_{L_q} \asymp m^{-g(\mathbf{r})} \tag{9.1}$$

for

$$1 < p < \infty, \quad 2 \leq q < \infty, \quad g(\mathbf{r}) > (1/p - 1/q)_+.$$

It is important to remark that the basis \mathcal{U}^d studied in [360] (see Section 7) realizes (9.1) for all \mathbf{r} . Thus, the orthonormal bases \mathcal{U}^d is optimal in the sense of order for m -term approximation

of classes of functions with mixed smoothness and also it is universal for m -term approximation of classes with anisotropic smoothness.

In this survey we discussed in detail the hyperbolic cross approximation, namely, approximation of periodic functions by the trigonometric polynomials with frequencies in the hyperbolic crosses. There is a natural analog of the trigonometric hyperbolic cross approximation in the wavelet approximation. In Sections 5, 7 we already discussed wavelet type systems \mathcal{U}^d . The general construction goes along the same lines. We briefly explain the construction from [71, 299], where it is called ‘‘Hyperbolic Wavelet Approximation’’. Let $\Psi = \{\psi_I\}$ be a system of univariate functions indexed by dyadic intervals. Define d -variate system $\Psi^d := \{\psi_{\mathbf{I}}\}$, $\mathbf{I} = I_1 \times \cdots \times I_d$, $\psi_{\mathbf{I}}(\mathbf{x}) := \prod_{j=1}^d \psi_{I_j}(x_j)$. Then the subspace

$$\Psi^d(n) := \text{span} \{ \psi_{\mathbf{I}} : |\mathbf{I}| \geq 2^{-n} \}$$

is an analog of the $\mathcal{T}(Q_n)$. The reader can find an introduction to the hyperbolic wavelet approximation in [71, 299].

9.3 Further generalizations

In this survey we discussed in detail some problems of linear and nonlinear approximation of functions with mixed smoothness. The term *mixed smoothness* means that in the definitions of the corresponding classes we either use the mixed derivative (classes **W**) or the mixed difference (classes **H** and **B**). Other classical scale of classes includes the Sobolev classes mentioned in the Introduction and the Hölder-Nikol’skii classes discussed above in Subsection 9.2 (for detailed study of approximation of these classes see [352], Chapter 2). The approximation properties and the techniques used for studying these classes are very different. There is an interesting circle of papers by D. Bazarkhanov [22] – [25] where he builds a unified theory, which covers both the collection of Sobolev, Hölder-Nikol’skii classes and classes with mixed smoothness. His approach is based on dividing the variable $\mathbf{x} = (x_1, \dots, x_d)$ into groups $\mathbf{x}^j := (x_{k_{j-1}+1}, \dots, x_{k_j})$, $j = 1, \dots, l$, $0 = k_0 < k_1 < \cdots < k_l = d$ and assuming that, roughly speaking, $f(\mathbf{x})$ as a function on each \mathbf{x}^j belongs to, say, a Sobolev class, and as a function on the variable $(\mathbf{x}^1, \dots, \mathbf{x}^l)$ has mixed smoothness.

We mostly confined ourselves to approximation in the Banach space L_q , $1 \leq q \leq \infty$, of functions from classes defined by a restriction on the mixed derivative or mixed difference in the Banach space L_p , $1 \leq p \leq \infty$. In our discussion parameters p and q are scalars. In some approximation problems it is natural to consider $L_{\mathbf{p}}$ spaces with vector $\mathbf{p} = (p_1, \dots, p_d)$ instead of a scalar p . For instance, considering general setting of approximation of classes of univariate functions (see [337])

$$W_p^g := \{f : f(x) = \int_{\Omega} g(x, y) \varphi(y) dy, \quad \|\varphi\|_{L_p(\Omega)} \leq 1\}$$

we immediately encounter the problem of approximation of the kernel $g(x, y)$ in vector norms. In particular, the operator norm of the integral operator with the kernel $g(x, y)$ from $L_p(\Omega)$ to $L_{\infty}(\Omega)$ is equal to $\|g(x, \cdot)\|_{L_p} \|g(\cdot, y)\|_{L_{\infty}}$. Dinh Dũng [84] – [89], Galeev [131] – [133] investigated embedding theorems, hyperbolic cross approximation and various widths for classes of functions with mixed smoothness defined in a vector $L_{\mathbf{p}}$ space in a vector $L_{\mathbf{q}}$ space. G.A. Akishev [2] – [3] conducted a detailed study of approximation of functions with mixed smoothness defined in a vector $L_{\mathbf{p}}$ space in general Lorentz-type spaces.

9.4 Direct and inverse theorems

In the univariate approximation theory classes with a given modulus of continuity (smoothness) are a natural generalization of classes H_∞^r . These classes were studied in detail from the point of view of direct and inverse theorems of approximation, which bear the names of Jackson and Bernstein. The reader can find the corresponding results in [73] and [352]. For instance it is well known that the following conditions are equivalent

$$E_n(f)_p \lesssim n^{-r}$$

$$\|f\|_{H_p^r} < \infty.$$

Clearly, researchers tried to prove direct and inverse theorems for approximation by the hyperbolic cross approximation. Direct theorems of that type for classes \mathbf{W} , \mathbf{H} and \mathbf{B} are discussed in Section 4. There are interesting results on the Jackson type theorems for the hyperbolic cross approximation. The first result in this direction was proved for approximation in the uniform metric in [338]. Further detailed study of approximation of classes of periodic functions of several variables with a given majorant of the mixed moduli of smoothness was conducted by Dinh Dũng [85, 86, 87, 88, 100], N.N. Pustovoitov [265] – [266] and by S.A. Stasyuk [307] – [308]. It was discovered in the early papers on the hyperbolic cross approximation that contrary to the univariate approximation by the trigonometric polynomials we cannot characterize classes of periodic functions of several variables with a given majorant of the mixed moduli of smoothness by the hyperbolic cross approximation, see also [286]. The problem of characterization of classes of functions with a given rate of decay of their best approximations by the hyperbolic cross polynomials was solved in [74, 92, 93]. As a result new concepts of a mixed modulus of smoothness were introduced in [74, 92].

9.5 Kolmogorov widths of the intersection of function classes

For $\mathbf{r} \in \mathbb{R}^d$, let the differential operator $D^{\mathbf{r}}$ be defined by $D^{\mathbf{r}}: f \mapsto (-i)^{|\mathbf{r}|_1} f^{(\mathbf{r})}$, where $f^{(\mathbf{r})}$ is the fractional derivative of order \mathbf{r} in the sense of Weil. For a nonempty finite set $A \subset \mathbb{N}_0^d$ and a nonzero sequence of numbers $(c_{\mathbf{r}})_{\mathbf{r} \in A}$, the polynomial $P(\mathbf{x}) := \sum_{\mathbf{r} \in A} c_{\mathbf{r}} \mathbf{x}^{\mathbf{r}}$ induce the differential operator

$$P(D) = \sum_{\mathbf{r} \in A} c_{\mathbf{r}} D^{\mathbf{r}}.$$

Set

$$U_2^{[P]} := \{f \in L_2 : \|P(D)(f)\|_2 \leq 1\}.$$

One of the most important problems in multivariate approximation is how to define smoothness function classes. In the early paper [6], Babenko suggested to define them as the functions from the finite intersection $\cap_{j=1}^J U^{[P_j]}$. There, he obtained a non-explicit upper bound of $d_n(\cap_{j=1}^J U_2^{[P_j]}, L_2)$ in terms of the eigenvalues of the operator $\sum_{j=1}^J P_j^* P_j$. The Sobolev class of mixed smoothness $\mathbf{W}_2^{\mathbf{r}}$ can be considered as a particular case of $U^{[P]}$. Tikhomirov and his school [84, 85, 86, 87, 88, 95, 97, 98, 135, 136, 138, 378, 379] considered smoothness functions as the functions from the intersections

$$\mathbf{W}_p^A := \cap_{\mathbf{r} \in A} \mathbf{W}_p^{\mathbf{r}}, \quad \mathbf{H}_p^A := \cap_{\mathbf{r} \in A} \mathbf{H}_p^{\mathbf{r}}, \quad \mathbf{B}_{p,\theta}^A := \cap_{\mathbf{r} \in A} \mathbf{B}_{p,\theta}^{\mathbf{r}}$$

for some (not necessarily finite) set $A \subset \mathbb{R}^d$. They investigated embedding theorems, hyperbolic cross approximation, various widths, entropy number and m -term approximation for these classes of functions on \mathbb{T}^d and \mathbb{R}^d .

As an illustration, we give a typical result in [84] on the Kolmogorov width $d_n(\mathbf{W}_p^A, L_p)$ for a nonempty finite set $A \subset [0, \infty)^d$ such that $\mathbf{0} \in A$ and $\max_{\mathbf{r} \in A} r_i > 0$ for every $i = 1, \dots, d$. Denote by $\text{conv}(A)$ the convex hull of A . We define two quantities characterizing the smoothness of \mathbf{W}_p^A .

$$r(A) := \min\{t > 0 : t(1, 1, \dots, 1) \in \text{conv}(A)\}$$

and $\nu(A) := d - \mu$ where μ is the dimension of the minimal face (extremal subset) of $\text{conv}(A)$ containing $r(1, 1, \dots, 1)$. These quantities were introduced by Dinh Dũng [82] in a dual form: $1/r(A)$ is defined as the optimal value of the problem

$$\text{maximize } |\mathbf{x}|_1, \quad \text{subject to } \mathbf{x} \in \mathbb{R}^d, \quad (\mathbf{r}, \mathbf{x}) \leq 1, \quad \forall \mathbf{r} \in A,$$

and $\nu(A)$ as the dimension of its solutions. In particular cases, we have $r(A) = r$, $\nu(A) = d - 1$ for the class $\mathbf{W}_p^r = \mathbf{W}_p^A$ with $A = \{r(1, 1, \dots, 1)\}$, and $r(A) = r_1$, $\nu(A) = \nu$, for the class $\mathbf{W}_p^{\mathbf{r}} = \mathbf{W}_p^A$ with $A = \{\mathbf{r}\}$, $0 < r_1 = \dots = r_\nu < r_{\nu+1} \leq r_{\nu+2} \leq \dots \leq r_d$ ($1 \leq \nu \leq d$). These examples tell us that the quantities $r(A)$ and $\nu(A)$ are indeed characteristics of the smoothness of \mathbf{W}_p^A .

Theorem 9.2. *For $1 < p < \infty$, we have*

$$d_n(\mathbf{W}_p^A, L_p) \asymp \left(\frac{\log^{\nu(A)-1} n}{n} \right)^{r(A)}, \quad n \in \mathbb{N}.$$

The problem of computing asymptotic orders of $d_n(U_2^{[P]}, L_2)$ in the general case when $W_2^{[P]}$ is compactly embedded into L_2 has been open for a long time; see, e.g., [352, Chapter 3] for details. It has been recently solved in [63] for a non-degenerate differential operator $P(D)$ (see there for a definition of non-degenerate differential operator). Here we give a generalization of this result which can be proven in a similar way.

Theorem 9.3. *Let P_j , $j = 1, \dots, J$ be polynomials with the power sets A_j and the coefficient sequences $(c_{\mathbf{r}}^j)_{\mathbf{r} \in A_j}$. Assume that the different operators $P_j(D)$ are non-degenerate, $0 \in A := \cap_{j=1}^J A_j$ and the intersection of A with the ray $\{\lambda e^j : \lambda > 0\}$ nonempty where e^j denotes the j th standard unit vector of \mathbb{R}^d . Then we have*

$$d_n(\cap_{j=1}^J U_2^{[P_j]}, L_2) \asymp \left(\frac{\log^{\nu(A)-1} n}{n} \right)^{r(A)}, \quad n \in \mathbb{N}. \quad (9.2)$$

Notice that in both Theorems 9.2 and 9.2, the asymptotic order of the Kolmogorov width is realized by the approximation by trigonometric polynomials with frequencies from the intersection of hyperbolic crosses corresponding to the set A . For related results, surveys and bibliography on embedding theorems, hyperbolic cross approximation, various widths, entropy number and m -term approximation of classes of multivariate periodic functions with several bounded fractional derivatives or bounded differences see [63, 84, 85, 86, 87, 88, 89, 95, 97, 98, 105, 106, 107, 132, 135, 136, 138, 221, 222, 223, 224, 378, 379].

9.6 Further s -numbers

It is well-known that every compact operator T in a Hilbert space H , i.e., $T : H \rightarrow H$ can be represented through its singular value decomposition (Schmidt expansion)

$$Tx = \sum_m s_m \langle x, u_m \rangle v_m, \quad x \in H,$$

where $(u_m)_m$ and $(v_m)_m$ are the eigenelements of T^*T and TT^* . Hence, we associate every compact operator a sequence $(s_m(T))_m$ of singular numbers.

The step in extending this concept to (quasi-)Banach spaces was done by Pietsch in 1974, see [259, 6.2.2] for more historical facts.

Definition 9.4. Let X, Z be quasi Banach-spaces and Y be a p -Banach space, let $S, T \in \mathcal{L}(X, Y)$ and $R \in \mathcal{L}(Y, Z)$. A mapping $s : T \rightarrow (s_m(T))_{m=0}^\infty$ with the following properties

$$(S1) \quad \|T\|_{\mathcal{L}(X, Y)} = s_0(T) \geq s_1(T) \geq \dots \geq 0,$$

(S2) for all $m_1, m_2 \in \mathbb{N}_0$ holds

$$s_{m_1+m_2}(R \circ S) \leq s_{m_1}(R)s_{m_2}(S),$$

(S3) for all $m_1, m_2 \in \mathbb{N}_0$ holds

$$s_{m_1+m_2}^p(S + T) \leq s_{m_1}^p(S) + s_{m_2}^p(T),$$

(S4) for all $m \in \mathbb{N}$ holds $s_m(id : \ell_2^m \rightarrow \ell_2^m) = 1, m = 1, 2, \dots$,

(S5) and $s_m(T) = 0$ whenever $\text{rank } T \leq m$,

is called *s-function*.

The definition of the widths studied in Section 4 (orthowidths, Kolmogorov widths, linear widths) can be extended in order to approximate general linear operators T instead of the identity/embedding operator. There are some issues related with this interpretation which are discussed in [259, 6.2.6] and [261, p. 30]. However, for Kolmogorov and approximation numbers the above axioms are satisfied and they form an *s-function*. In other words, they represent special sequences of *s-numbers*, $d_m(T)$ and $a_m(T)$. We use the usual notation $a_m(T)$ (instead of $\lambda_m(T)$) in order to avoid confusion with the sequence of eigenvalues. A simple consequence of (S5) is the fact that the approximation numbers form the largest sequence of *s-numbers*. In fact, for any operator with rank less or equal to m it holds

$$s_m(T) \leq s_m(S) + \|T - S\| = \|T - S\|$$

and therefore $s_m(T) \leq a_m(T)$. As a direct implication we see that the orthowidths (discussed in Section 4) can not be interpreted as *s-numbers*. In fact, in some situations they are asymptotically larger than approximation numbers (linear widths), see Section 4.

Let us emphasize that the sequence of dyadic entropy numbers $\epsilon_m(T)$ (studied in Chapter 6) does not give an *s-functions* since it does not satisfy (S5). However, there are many more interesting examples of *s-numbers*. Let us first discuss Gelfand numbers/widths.

The *Gelfand widths* have been introduced by Tikhomirov in 1965 (Gelfand just proposed them)

$$c_m(\mathbf{F}, X) := \inf_{\substack{A: X \rightarrow \mathbb{R}^m \\ \text{linear}}} \sup_{f \in \mathbf{F} \cap \ker A} \|f\|_X. \quad (9.3)$$

They can be interpreted as *s-numbers*, the so-called Gelfand numbers $c_m(T)$ ([259, 6.2.3.3]) and satisfy a useful duality relation. For any compact operator $T : X \rightarrow Y$ it holds

$$c_m(T^* : Y' \rightarrow X') = d_m(T : X \rightarrow Y) \quad \text{and} \quad d_m(T^* : Y' \rightarrow X') = c_m(T : X \rightarrow Y),$$

where T^* denotes the dual operator to T and X', Y' the dual spaces to X and Y . As a special case we obtain the relation

$$d_m(\mathbf{W}_p^r, L_q) = c_m(\mathbf{W}_{q'}^r, L_{p'}) \quad , \quad m \in \mathbb{N}.$$

if $1 < p, q < \infty$. “Dualizing” Figure 4 gives the following order (compared to linear widths).

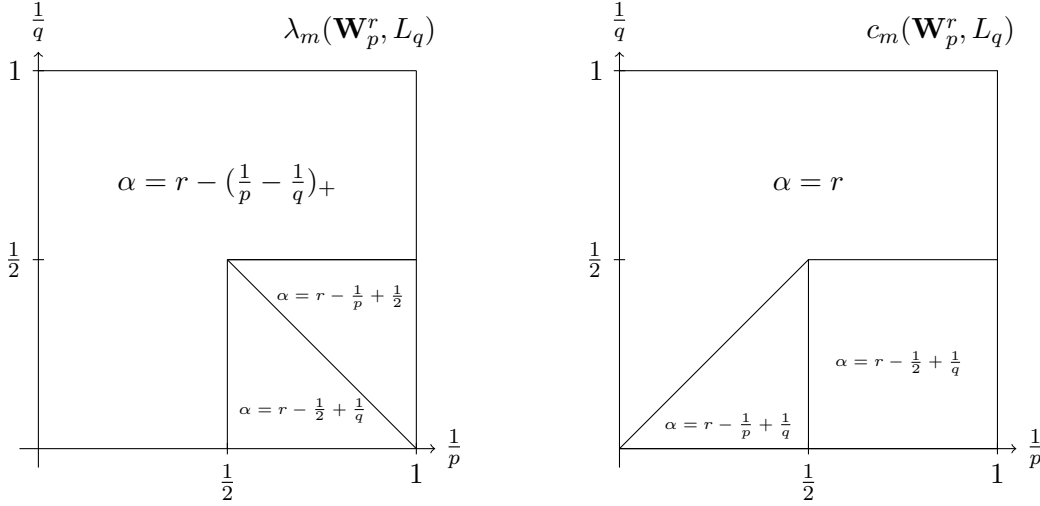


Figure 21: Comparison of $\lambda_m(\mathbf{W}_p^r, L_q)$ and $c_m(\mathbf{W}_p^r, L_q)$, rate $(m^{-1}(\log m)^{d-1})^\alpha$

The Gelfand widths can be interpreted from an algorithmic point of view. Suppose, one is interested in reconstructing an object (function) from m linear samples. The mapping A serves as measurement map and is linear, whereas the reconstruction can be a nonlinear operator using the measurement vector from \mathbb{R}^m . In that sense they directly relate to the novel field of compressed sensing, see [125, Chapt. 10], [368, Chapt. 5], and the recent paper [124]. An optimal measurement map minimizes (9.3) and measures the maximal “distance” of two instances which provide the same output (by A). Hence, any reconstruction map can not distinguish between those two instances and the reconstruction error would relate to their distance. In the picture there are parameter regions where the linear widths are asymptotically strictly larger than the Gelfand widths. In this region a general (possibly nonlinear) operator/algorithm (that uses linear information) might beat the linear algorithm which is behind the linear widths. Such an algorithm can probably come out of an optimization like in learning theory or compressed sensing.

In addition, some of the s -numbers are useful for upper bounding the sequence of eigenvalues of compact operators (Weyl inequalities, [259, 6.4.2]). For this reason authors studied the order of Weyl numbers $x_m(T)$ also in the context of function spaces with bounded mixed difference/derivative, see Nguyen, Sickel [236, 237, 238]. In contrast to the Weyl numbers, which represent special s -numbers, the Bernstein number $b_m(T)$ fail to satisfy (S2, S3). However, Bernstein numbers are important since they often serve as lower bounds for Kolmogorov and Gelfand numbers, and especially, as lower bounds for the error analysis of Monte-Carlo algorithms, see the recent work [211].

V.K. Nguyen shows in [238, Lem. 3.3] the interesting relation

$$b_m(T) \leq 2\sqrt{2}\epsilon_m(T) \quad , \quad m \in \mathbb{N}.$$

Let us shortly write $b_m(X, Y)$ and $x_m(X, Y)$ when considering the identity mapping $b_m(I)$ and $x_m(I)$ for $I : X \rightarrow Y$. Together with an abstract relation between Bernstein and Weyl numbers

[260] this leads to the sharp relation

$$b_m(\mathbf{W}_p^r, L_q) \asymp \min\{x_m(\mathbf{W}_p^r, L_q), \epsilon_m(\mathbf{W}_p^r, L_q)\} \quad , \quad m \in \mathbb{N}.$$

The results on Bernstein and Weyl numbers can be illustrated in the following diagrams.

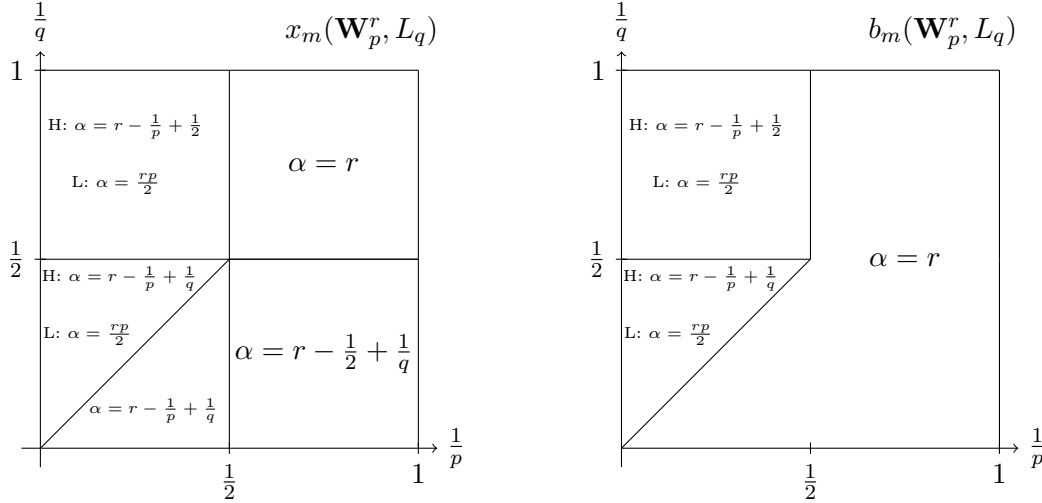


Figure 22: Comparison of $x_m(\mathbf{W}_p^r, L_q)$ and $b_m(\mathbf{W}_p^r, L_q)$, rate $(m^{-1}(\log m)^{d-1})^\alpha$

Here, H refers to the parameter domain of “high smoothness”, $r > \frac{1/\max\{2,q\}-1/p}{p/2-1}$ and L refers to “low smoothness”, that is $r < \frac{1/\max\{2,q\}-1/p}{p/2-1}$.

9.7 The quasi-Banach situation

Complements to Section 3: Mixed differences

By replacing the moduli of continuity (3.6) by more regular variants like integral means (3.11) of differences we can extend the characterization in Lemma 3.10 to all $0 < p \leq \infty$ and $r > (1/p - 1)_+$.

For $\mathbf{s} \in \mathbb{N}_0^d$ we put $2^{-\mathbf{s}} := (2^{-s_1}, \dots, 2^{-s_d})$ and $e(\mathbf{s}) = \{i : j_i \neq 0\}$. We have the following characterization, see [239] and [390].

Theorem 9.5. *Let $0 < p, \theta \leq \infty$ and $s > \sigma_p$. Let further $m \in \mathbb{N}$ be a natural number with $m > s$. Then*

$$\|f\|_{\mathbf{B}_{p,\theta}^s} \asymp \left(\sum_{\mathbf{s} \in \mathbb{N}_0^d} 2^{s|\mathbf{s}|_1 \theta} \|\mathcal{R}_m^{e(\mathbf{s})}(f, 2^{-\mathbf{s}}, \cdot)\|_p^\theta \right)^{1/\theta}.$$

In case $\theta = \infty$ the sum above is replaced by the supremum over \mathbf{s} .

Complements to Section 5: Sampling representations

An extension of Propositions 5.3, 5.4 to all parameters $0 < p, \theta \leq \infty$ and $r > \max\{0, 1/p - 1\}$ is possible but not straight forward. One can not expect, that the de la Vallée Poussin sampling representation (5.18) works for all $0 < p < 1$. In order to get a sampling representation for arbitrary small $0 < p < 1$ we have to use different fundamental interpolants in (5.13). Those have to provide a better decay than (2.8) on the “time side” or, in different words, the

Fourier transform of the kernel has to be much smoother than just piecewise linear. This can be arranged, see [48], by an iterated convolution of dilates of the characteristic function of the interval $[-1/2, 1/2]$, a process related to the construction of the so-called *up*-function, see [284]. However, when dealing with sampling on Smolyak grids it turned out, see [48], that the integrability in the target space L_q determines the appropriate sampling kernel. In fact, when approximating functions from the class $\mathbf{B}_{p,\theta}^r$, with $p < 1$ in L_2 the de la Vallée Poussin-Smolyak operator (5.24) still does the job.

The Faber-Schauder and B-spline quasi-interpolation relations in Propositions 5.19, 5.18, 5.24,(i) for the periodic space $\mathbf{B}_{p,\theta}^r$ as well as its non-periodic counterpart can be extended to $0 < p, \theta \leq \infty$. In contrast to the de la Vallée Poussin kernels the hat functions are perfectly localized. In other words, we have sufficient decay on the time side. However, the smoothness of the piecewise linear hat functions is limited. Therefore, one can not expect to decompose arbitrary smooth function over the Faber-Schauder basis. This explains the natural restriction $1/p < r < 2$ in Proposition (5.18) and hence we can only go “down” to $p > 1/2$ and only get a maximal rate of convergence (worst-case error) bounded by 2. Using B-splines one can achieve higher order convergence by adapting the order of the B-spline also to p (and not just on r). Note that the restriction in Proposition 5.24 reads as $1/p < r < 2\ell$ if $p < 1$. In particular this means that we have to choose 2ℓ larger than $1/p$, a similar effect as described in the previous paragraph.

Complements to Section 5: Sampling widths

The method described in the proof of Theorem 5.9 also works for $p, \theta, q < 1$ once we have the corresponding sampling representations. For this issue we refer [99] and the recent Preprint [48]. Here we obtain the following results.

Theorem 9.6. *Let $0 < p, q, \theta \leq \infty$, and $r > 1/p$. Then we have.*

(i) *For $p \geq q$ and $\theta \leq 1$,*

$$\varrho_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp (m^{-1} \log^{d-1} m)^r, \quad \begin{cases} 2 \leq q < p < \infty, \\ 1 < p = q \leq \infty. \end{cases}$$

(ii) *For $1 < p < q < \infty$,*

$$\varrho_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp (m^{-1} \log^{d-1} m)^{r-1/p+1/q} (\log^{d-1} m)^{(1/q-1/\theta)_+}, \quad \begin{cases} 2 \leq p, 2 \leq \theta \leq q, \\ q \leq 2. \end{cases}$$

Complements to Section 8: Numerical integration

Theorem 9.7. (ii) *For each $0 < p, \theta \leq \infty$ and $r > 1/p$, we have*

$$\Phi(a, A, \psi)(\mathbf{B}_{p,\theta}^r) \asymp a^{-dr+d(1/p-1)_+} (\log a)^{(d-1)(1-1/\theta)_+}, \quad a > 1.$$

(ii) *For each $0 < p < \infty$ and $0 < \theta \leq 1$, we have*

$$\Phi(a, A, \psi)(\mathbf{B}_{p,\theta}^{1/p}) \asymp a^{-d/\max\{p,1\}}, \quad a > 1.$$

Analogous estimates hold true for the Fibonacci-cubature formulas.

Nonlinear approximation of non-compact embeddings

In contrast to the study of s -numbers it makes also sense to study non-compact (but continuous) embeddings in connection with best m -term approximation. This has been done recently in [164]. Let us first clarify the precise embedding situation. A proof may be found in [396].

Lemma 9.8. *Let $1 < q < \infty$ and $r > 0$.*

(i) *Let $1 < p < \infty$. Then*

$$\mathbf{W}_p^r \hookrightarrow L_q$$

if and only if $r \geq 1/p - 1/q$.

(ii) *Let $0 < p < \infty$. Then*

$$\mathbf{B}_{p,p}^r \hookrightarrow L_q$$

if and only if $r \geq 1/p - 1/q$.

(iii) *The embeddings are compact if and only if $r > 1/p - 1/q$. The embeddings (i) and (ii) also hold true for spaces $\mathbf{B}_{p,p}^r(\mathbb{R}^d)$ and $\mathbf{W}_p^r(\mathbb{R}^d)$ if $p \leq q$ and $r \geq 1/p - 1/q$ but they are never compact.*

The compact embeddings $r > 1/p - 1/q$ have been discussed above. Now we are interested in $r = 1/p - 1/q$. Again we use a wavelet basis of sufficiently smooth wavelets as dictionary Φ .

Theorem 9.9. (i) *Let $0 < p \leq \max\{p, 1\} < q < \infty$ and $r = 1/p - 1/q$. Then*

$$\sigma_m(\mathbf{B}_{p,p}^r, \Phi)_q \asymp m^{-r} (\log m)^{(d-1)(r-1/p+1/2)+}.$$

(ii) *Let $1 < p \leq 2 \leq q < \infty$ and $r = 1/p - 1/q > 0$. Then*

$$\sigma_m(\mathbf{W}_p^r, \Phi)_q \asymp m^{-r} (\log m)^{(d-1)r}.$$

Remark 9.10. *In the case $q = 2$ we obtain with $r = 1/p - 1/2$*

$$\sigma_m(\mathbf{B}_{p,p}^r, \Phi)_2 \asymp m^{-r}. \quad (9.4)$$

This has been already observed in [299], see also [245]. The latter reference dealt with tensor products of univariate Besov spaces $B_{p,p}^r$ which have been recently identified as spaces with dominating mixed smoothness also in the case $p < 1$, see [299].

We can say even more in the situation $q = 2$ for spaces on \mathbb{R}^d . For $s > 0$ and p as above we consider the nonlinear approximation space $\mathbf{A}_p^s(L_2(\mathbb{R}^d))$, see e.g. [69], as the collection of all $f \in L_2(\mathbb{R}^d)$ such that

$$\|f\|_{\mathbf{A}_p^s(L_2(\mathbb{R}^d))} := \|f\|_2 + \left(\sum_{m=1}^{\infty} \frac{1}{m} [m^s \sigma_m(f, \Phi)_2]^p \right)^{1/p} < \infty.$$

Mainly as a corollary of the characterization of $\mathbf{A}_p^{\frac{1}{p}-\frac{1}{2}}(L_2(\mathbb{R}^d))$, see e.g. Pietsch [258], DeVore [69, Thm. 4] or [368, Sect. 1.8], we obtain the following identification.

Theorem 9.11. *Let $0 < p < 2$. Then we have*

$$\mathbf{A}_p^{\frac{1}{p}-\frac{1}{2}}(L_2(\mathbb{R}^d)) = \mathbf{B}_{p,p}^{1/p-1/2}(\mathbb{R}^d)$$

in the sense of equivalent quasi-norms.

9.8 Sampling along lattices

In Section 5 (Subsection 5.4) we encountered several situation where Smolyak’s algorithm (sparse grids) represent the optimal sampling algorithm with respect to ϱ_m . However, in case $q \leq p$ an optimal sampling algorithm is only known in case $d = 2$ and $p = q = \infty$. The natural question arises whether there are other discrete point set constructions and corresponding sampling algorithms which might behave better than sparse grids or perform even optimal. Here one could think about the big variety of point set constructions which are used for numerical integration based on number theoretic constructions, see [80] or Section 8 above. This question has been considered by several authors in the past. We refer to [333], [213], [214], [189], [187], [190], [191]. It turned out that the proposed methods provide certain numerical advantages since the lattice structure allows for using the classical Fast Fourier Transform for the evaluation and reconstruction of trigonometric polynomials. Moreover, in contrast to sparse grid discretization techniques the use of oversampled lattice rules provides better numerical stability properties [189], [188]. However, when expressing the error in terms of the used function values (in the sense of ϱ_m) it turns out, see the recent paper [50], that one only can expect half the rate of convergence of a sparse grid method, i.e.

$$\varrho_m^{\text{latt}_1}(\mathbf{W}_{2,r}^r, L_2) \gtrsim m^{-r/2}.$$

The quantities on the left-hand side are similarly defined as ϱ_m above with the difference that we take the inf over all sampling algorithms taking function values along rank-1 lattices with m points. Surprisingly, such methods never perform optimal with respect to the “information complexity”. However, as already mentioned in Subsection 5.3, the runtime of such an algorithm can be accelerated significantly by exploiting the lattice structure.

A related framework is studied in the papers [10], [251]. The authors deal with sampling recovery of the class or rank-1 tensors

$$F_{M,d}^r := \left\{ f = \bigotimes_{i=1}^d f_i : \|f_i\|_\infty \leq 1, \|f_i^{(r)}\|_\infty \leq M \right\}$$

which represents a subclass of the unit ball in the space \mathbf{W}_∞^r if $M = 1$. The authors in [10] use a low discrepancy point set (Halton points, see Section 8) to propose a sampling recovery algorithm with N points and L_∞ -error $\lesssim C(r, d)N^{-r}$ with constant $C(d, r)$ scaling like d^{dr} . The question after the curse of dimensionality arises. By proposing a Monte-Carlo sampling recovery algorithm the authors in [251] prove that the curse of dimensionality is present (in the randomized setting) if and only if $M \geq 2^r r!$. In case $M < 2^r r!$ the complexity is only polynomial in the dimension.

9.9 Sampling recovery in energy norm

There is a large class of solutions of the electronic Schrödinger equation in quantum chemistry, which belong to Sobolev spaces with mixed regularity and, moreover, possess some additional Sobolev isotropic smoothness properties, see Yserentant’s recent lecture notes [406] and the references therein. This type of regularity is precisely expressed by the spaces $\mathbf{W}_p^{r,\beta}$ for $p = 2$ introduced in [153]. Here, the parameter r reflects the smoothness in the dominating mixed sense and the parameter β reflects the smoothness in the isotropic sense. In [44, 45, 152, 151, 153, 293], the authors used Galerkin methods for the $W_2^1(\mathbb{T}^d)$ -approximation of the solution of general elliptic variational problems. In particular, by use of tensor-product biorthogonal wavelet bases, the authors of [153] constructed so-called optimized sparse grid

subspaces for finite element approximations of the solution having $\mathbf{W}_2^{r,\beta}$ -regularity, whereas the approximation error is measured in the energy norm of isotropic Sobolev space W_2^γ . For non-periodic functions of mixed smoothness r from \mathbf{W}_2^r , linear sampling algorithms on sparse grids have been investigated by Bungartz and Griebel [45] employing hierarchical Lagrangian polynomials multilevel basis and measuring the approximation error in energy W_2^1 -norm.

The problem of sampling recovery on sparse grids in energy norm W_q^γ of functions from classes $\mathbf{W}_p^{r,\beta}$ and $\mathbf{B}_{p,\theta}^{r,\beta}$ has been investigated in [47, 100]. The space $\mathbf{B}_{p,\theta}^{r,\beta}$ is a “hybrid” of the space $\mathbf{B}_{p,\theta}^r$ of mixed smoothness r and the classical isotropic Besov space $B_{p,\theta}^\beta$ of smoothness β . The space $\mathbf{B}_{p,\theta}^{r,\beta}$ can be seen as a Besov type generalization of the space $\mathbf{W}_2^{r,\beta} = \mathbf{B}_{2,2}^{r,\beta}$. For the definition of these spaces, see [47, 100]. It turns out that for this sampling recovery, we can achieve the right order of the sampling widths.

Theorem 9.12. *Let $r, \beta, \gamma \in \mathbb{R}$ such that $\min\{r, r + \beta\} > 1/2$, $\gamma \geq 0$ and $0 < \gamma - \beta < r$. Then we have*

$$\varrho_m(\mathbf{W}_2^{r,\beta}, W_2^\gamma(\mathbb{T}^d)) \asymp m^{-(r-\gamma+\beta)}.$$

Theorem 9.13. *Let $0 < p, \theta \leq \infty$, $1 < q < \infty$, $r > 0$, $\gamma \geq 0$ and $\beta \in \mathbb{R}$, $\beta \neq \gamma$. Assume that there hold the conditions $\min(r, r + \beta) > 1/p$ and*

$$r > \begin{cases} (\gamma - \beta)/d, & \beta > \gamma, \\ \gamma - \beta, & \beta < \gamma. \end{cases}$$

Then we have

$$\varrho_m(\mathbf{B}_{p,\theta}^{r,\beta}, W_q^\gamma(\mathbb{I}^d)) \asymp \begin{cases} m^{-r-(\beta-\gamma)/d+(1/p-1/q)_+}, & \beta > \gamma, \\ m^{-r-\beta+\gamma+(1/p-1/q)_+}, & \beta < \gamma. \end{cases}$$

Theorem 9.12 has been proven in [47] and Theorem 9.13 in [100]. Special sparse grids are constructed for sampling recovery of $\mathbf{W}_p^{r,\beta}$ and $\mathbf{B}_{p,\theta}^{r,\beta}$. They have much smaller number of sample points than the corresponding standard full grids and Smolyak grids, but give the same error of the sampling recovery on the both latter ones. The construction of asymptotically optimal linear sampling algorithms is essentially based on quasi-interpolation representations by Dirichlet kernel series or B-spline series of functions from $\mathbf{W}_p^{r,\beta}$ and $\mathbf{B}_{p,\theta}^{r,\beta}$ with a discrete equivalent quasi-norm in terms of the coefficient function-valued functionals of this series. For details and more results on this topic see [47, 100].

9.10 Continuous algorithms in m -term approximation and nonlinear widths

Notice that if X is separable Banach space and the dictionary \mathcal{D} is dense in the unit ball of X , then the best m -term approximation of f with respect to \mathcal{D} vanishes, i.e., $\sigma_m(f, \mathcal{D})_X = 0$ for any $f \in X$. On the other hand, for almost all well-known dictionaries with good approximation properties the quantity $\sigma_m(\mathbf{F}, \mathcal{D})_X$ has reasonable lower bounds for well-known classes of functions \mathbf{F} having a common smoothness. Therefore, the first problem which actually arises, is to impose additional conditions on \mathcal{D} and/or methods of m -term approximation. In [94, 195] to obtain the right order of $\sigma_m(\mathbf{F}, \mathcal{D})_X$ for certain function classes the dictionary \mathcal{D} is required to satisfy some “minimal properties”.

Another approach to dealing with this problem is to impose continuity assumptions on methods of m -term approximation by the elements of $\Sigma_m(\mathcal{D})$ which is the set of all linear combinations g of the form $g = \sum_{j=1}^m a_j g_j$, $g_j \in \mathcal{D}$. This does not weaken the rate of the m -term approximation for many well-known dictionaries and function classes. The continuity

assumptions on approximation methods certainly lead to various notions of nonlinear width, in particular, the classical Alexandroff width and the nonlinear manifold width [70, 231]. Thus, the classical Alexandroff width $a_m(\mathbf{F}, X)$ characterizes the best approximation of \mathbf{F} by continuous methods as mappings from \mathbf{F} to a complex of topological dimension $\leq n$ (see, e.g., [379] for an exact definition). Let us recall one of the notions of nonlinear widths based on continuous methods of m -term approximation suggested in [95].

A continuous method of m -term approximation with regard to the dictionary \mathcal{D} is called a continuous mapping from X into $\Sigma_m(\mathcal{D})$. Denote by \mathbb{F} the set of all dictionaries \mathcal{D} in X such that the intersection of \mathcal{D} with any finite dimensional linear subspace is a finite set, and by $C(X, \Sigma_m(\mathcal{D}))$ the set of all continuous mappings S from X into $\Sigma_m(\mathcal{D})$. The restriction with continuous methods of m -term approximation and dictionaries $\mathcal{D} \in \mathbb{F}$ leads to the notion of nonlinear width:

$$\alpha_m(\mathbf{F}, X) := \inf_{\mathcal{D} \in \mathbb{F}} \inf_{S \in C(X, \Sigma_m(\mathcal{D}))} \sup_{x \in \mathbf{F}} \|f - S(f)\|_X.$$

Theorem 9.14. *Let $1 < p, q < \infty$ and $2 \leq \theta \leq \infty$ and $r > \max(0, 1/p - 1/q, 1/p - 1/2)$. Then we have*

$$a_m(\mathbf{W}_p^r, L_q) \asymp \alpha_m(\mathbf{W}_p^r, L_q) \asymp m^{-r} (\log m)^{(d-1)r},$$

and

$$a_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp \alpha_m(\mathbf{B}_{p,\theta}^r, L_q) \asymp m^{-r} (\log m)^{(d-1)(r+1/2-1/\theta)}.$$

This theorem was proven in [95]. A different concept of non-linear width based on pseudo-dimension related to Learning Theory, but without continuity assumptions was suggested in [267, 268] and investigated for function classes of mixed smoothness in [96, 97]. For other notions of nonlinear widths, related results, surveys and bibliography, see [95, 96, 97, 110, 69, 70, 72, 195, 231, 267, 268, 311, 377, 379].

10 High-dimensional approximation

We explained in Subsections 9.1 and 9.2 that classes with mixed smoothness play a central role among the classes of functions with finite smoothness. A typical error bound for numerical integration, sampling, and hyperbolic cross approximation of, say, \mathbf{W}_p^r in L_p is of the form $m^{-r}(\log m)^{(d-1)\xi}$, where $\xi > 0$ may depend on r and p . The classical setting of the problem of the error behavior asks for dependence on m , when other parameters r, p, d are fixed. In this survey we mostly discuss this setting. Then the factor $(\log m)^{(d-1)\xi}$ plays a secondary role compared to the factor m^{-r} . However, as we pointed out in the Introduction, the problems with really large d attract a lot of attention. In this case, for reasonably small m the factor $(\log m)^{(d-1)\xi}$ becomes a dominating one and makes it impossible to apply a model of classes with finite smoothness for practical applications. It is a very important and difficult problem of contemporary numerical analysis. One of the promising ways to resolve this problem is to use classes of functions with special structure instead of smoothness classes. This approach is based on the concept of *sparsity* (with respect to a dictionary) and is widely used in compressed sensing and in greedy approximation. For instance, sparsity with respect to a dictionary with tensor product structure is a popular structural assumption (see Section 6).

10.1 Anisotropic mixed smoothness

We point out that classes with mixed smoothness also have a potential for applications in high-dimensional approximation. To illustrate this point we begin with the anisotropic version of classes \mathbf{W} , \mathbf{H} , \mathbf{B} . So far we discussed isotropic classes of functions with mixed smoothness. Isotropic means that all variables play the same role in the definition of our smoothness classes. In the hyperbolic cross approximation theory anisotropic classes of functions with mixed smoothness are of interest and importance. The framework described below goes back to the work of Mityagin [233, pp. 397, 409] in 1962 and Telyakovskii [315, p. 438] in 1964 and has been later used by several authors from the former Soviet Union, see [340, pp. 32, 36, 72] (English version) for further historical comments.

We give the corresponding definitions. Let $\mathbf{r} = (r_1, \dots, r_d)$ be such that $0 < r_1 = r_2 = \dots = r_\nu < r_{\nu+1} \leq r_{\nu+2} \leq \dots \leq r_d$ with $1 \leq \nu \leq d$. For $\mathbf{x} = (x_1, \dots, x_d)$ denote

$$F_{\mathbf{r}}(\mathbf{x}) := \prod_{j=1}^d F_{r_j}(x_j)$$

and

$$\mathbf{W}_p^{\mathbf{r}} := \{f : f = \varphi * F_{\mathbf{r}}, \quad \|\varphi\|_p \leq 1\}.$$

We now proceed to classes $\mathbf{H}_q^{\mathbf{r}}$ and $\mathbf{B}_{q,\theta}^{\mathbf{r}}$. Define

$$\|f\|_{\mathbf{H}_q^{\mathbf{r}}} := \sup_{\mathbf{s}} \|\delta_{\mathbf{s}}(f)\|_q 2^{(\mathbf{r}, \mathbf{s})},$$

and for $1 \leq \theta < \infty$ define

$$\|f\|_{\mathbf{B}_{q,\theta}^{\mathbf{r}}} := \left(\sum_{\mathbf{s}} \left(\|\delta_{\mathbf{s}}(f)\|_q 2^{(\mathbf{r}, \mathbf{s})} \right)^\theta \right)^{1/\theta}.$$

We will write $\mathbf{B}_{q,\infty}^{\mathbf{r}} := \mathbf{H}_q^{\mathbf{r}}$. Denote the corresponding unit ball

$$\mathbf{B}_{q,\theta}^{\mathbf{r}} := \{f : \|f\|_{\mathbf{B}_{q,\theta}^{\mathbf{r}}} \leq 1\}.$$

It is known that in many problems of estimating asymptotic characteristics the anisotropic classes of functions of d variables with mixed smoothness behave in the same way as isotropic classes of functions of ν variables (see, for instance, [340, 88, 316]). It is clear that the above remark holds for the lower bounds. To prove it for the upper bounds one needs to adapt the hyperbolic cross approximation to the smoothness of the class. We pay more samples in those directions where the smoothness is small. This results in an anisotropic hyperbolic cross.

What concerns numerical integration, the studied methods like the Fibonacci and Frolov cubature formulas need not to be adapted. These methods are able to “detect” the anisotropy themselves.

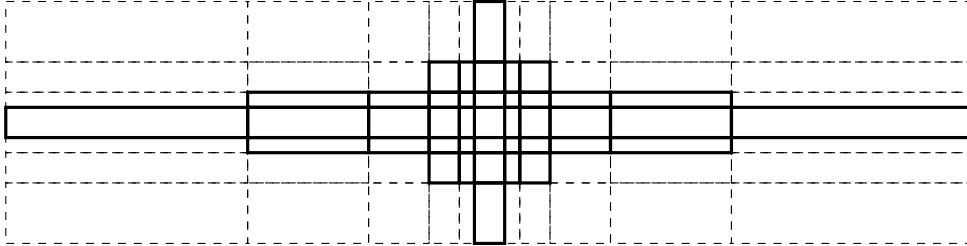


Figure 23: Anisotropic hyperbolic cross for $\mathbf{r} = (1, 2)$ in $d = 2$

The consequence is that the logarithmic factor $(\log m)^{(d-1)\xi}$ is replaced by the factor $(\log m)^{(\nu-1)\xi}$. Thus, if we use a model of anisotropic class with mixed smoothness \mathbf{r} with small ν , the factor $(\log m)^{(\nu-1)\xi}$ is not a problem anymore. In this case it is important to study dependence on d of the constants. In the recent paper [104], it has been shown that with a fixed ν and some moderate conditions on the anisotropic mixed smoothness \mathbf{r} the rate of the hyperbolic cross approximation does not depend on the dimension d , when d may be very large or even infinite (see also Subsection 10.4).

10.2 Explicit constants and preasymptotics

As already mention there is an increasing interest in approximation problems where the dimension d is large or even huge. Since then people were not only interested in getting the right order of the respective approximation error, also the dependence of the error bounds on the underlying dimension d became crucial. In case of high dimensions, the traditional estimate (1.2) becomes problematic. Let us illustrate this issue in the following figure. Fixing d and r the function $f_d(t) := t^{-r} (\log t)^{r(d-1)}$ is increasing on $[1, e^{d-1}]$ and decreasing on $[e^{d-1}, \infty)$.

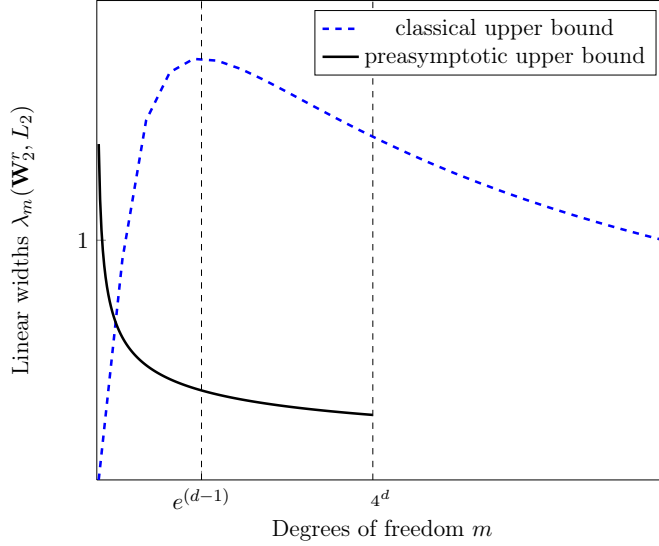


Figure 24: Asymptotics vs. preasymptotics

Hence, its maximum on $[1, \infty)$ is $\max_{t \geq 1} f_d(t) = f_d(e^{d-1}) = \left(\frac{d-1}{e}\right)^{r(d-1)}$, which increases super-exponentially in d . That means, for large d we have to wait “exponentially long” until the sequence $n^{-r}(\ln n)^{(d-1)r}$ decays, and even longer until it becomes less than one. This observation indicates the so-called “curse of dimensionality”, a phrase originally coined by Bellmann [28] in 1957. For a mathematically precise notion see [253, p. 1]. How to avoid that the logarithm increases exponentially in d has been discussed in Subsection 10.1. This could be one approach for making the problem tractable in high dimensions. In any case it is important to control the behavior of the constants $c(r, d)$ and $C(r, d)$ in d and, in a second step, to establish *preasymptotic estimates* for small m .

The first result on upper bounds in this direction (to the authors’ knowledge) has been given by Wasilkowski, Woźniakowski [401, Sect. 4.1]. There the authors studied Smolyak’s algorithm in a general framework (similar as done in Section 4.2) and obtained upper bounds for the error based on the behavior of the participating univariate operators.

Linear and Kolmogorov widths of Sobolev classes of a mixed smoothness in $L_2(\mathbb{T}^d)$ -norm and so-called energy norms in the high-dimensional setting have been first investigated by Dinh Dũng and Ullrich in [108] where it is stressed to treat the d -dependent upper bound and lower bound of these widths together. Of course, the specific definition of the space is now of particular relevance. Here, the class \mathbf{W}_2^r is defined as the set of function in $L_2(\mathbb{T}^d)$ for which the right hand side of (3.5) is finite. In particular, the following almost precise upper and lower bounds explicit in the dimension d have been proven in [108] (see also Theorem 10.4 below).

Theorem 10.1. *For any $r > 0$ and $m \geq 2^d$*

$$\frac{1 + \log e}{4^r(d-1)^{r(d-1)}} \left(\frac{m}{(\log m)^{d-1}} \right)^{-r} \leq d_m(\mathbf{W}_2^r, L_2(\mathbb{T}^d)) \leq 4^r \left(\frac{2e}{d-1} \right)^{r(d-1)} \left(\frac{m}{(\log m)^{d-1}} \right)^{-r}.$$

One observes a super-exponential decay of the constants. Other results have been obtained recently by Kühn, Sickel, Ullrich [210]. In contrast to Theorem (10.1) the authors considered the space \mathbf{W}_2^r to be specifically normed as follows

$$\|f\|_{\mathbf{W}_2^r}^2 := \sum_{\mathbf{k} \in \mathbb{Z}^d} |\hat{f}(\mathbf{k})|^2 \prod_{j=1}^d (1 + |k_j|^2)^r. \quad (10.1)$$

It has been shown in [210] that in this situation the “asymptotic constant” behaves exactly as follows

$$\lim_{m \rightarrow \infty} \frac{m^r}{(\log m)^{(d-1)r}} \cdot \lambda_m(\mathbf{W}_2^r, L_2) = \left[\frac{2^d}{(d-1)!} \right]^r. \quad (10.2)$$

This result is surprising from several points of view. First, the limit exists, second one can compute it explicitly and third, the number on the right-hand side decays exponentially in d . The next step is to ask for estimates of type (1.2) with precise given constants in some range $m \geq m_0$. There are several results in this direction, see for instance [45, Thm. 3.8], [108], [293, Thm. 5.1], where the super-exponential decay of the constant has been already observed in different periodic and non-periodic settings. For a thorough discussion and comparison of the mentioned results we refer to [210, Section 4.5]. To be consistent with the setting described in (10.1) let us state the following result from [210].

Theorem 10.2. *Let $r > 0$ and $d \in \mathbb{N}$. Then we have*

$$\lambda_m(\mathbf{W}_2^r, L_2) \leq \left[\frac{(3 \cdot \sqrt{2})^d}{(d-1)!} \right]^r \frac{(\ln m)^{(d-1)r}}{m^r}, \quad \text{if } m \geq 27^d$$

and

$$\lambda_m(\mathbf{W}_2^r, L_2) \geq \left[\frac{3}{d!} \left(\frac{2}{2 + \ln 12} \right)^d \right]^r \frac{(\ln m)^{(d-1)r}}{m^r}, \quad \text{if } m > (12 e^2)^d.$$

Note, that the constants decay super-exponentially in d as expected from (10.2). However, if d is large we have to “wait” very long until that happens. Hence, the next question is what happens in the preasymptotical range, say for m less than 2^d . Here, we get the bound below in Theorem 10.3 (see [210, Thm. 4.17] and Figure 24 above). For similar results in more general classes as well as a non-periodic counterpart, see the longer arXiv version of [56].

Theorem 10.3. *Let $d \geq 2$, and $r > 0$. Then for any $2 \leq m \leq \frac{d}{2} 4^d$ we have the upper estimate*

$$\lambda_m(\mathbf{W}_2^r, L_2) \leq \left(\frac{e^2}{m} \right)^{\frac{r}{4+2 \log_2 d}}.$$

Note, that there is no hidden constant in the upper bound. This type of error decay reflects “quasi-polynomial” tractability, a notion recently introduced by Gnewuch, Woźniakowski [148]. The result in Theorem 10.3 is based on a refined estimate for the cardinality of the hyperbolic cross

$$\tilde{\Gamma}(N, d) := \left\{ k \in \mathbb{Z}^d : \prod_{j=1}^d (1 + |k_j|) \leq N \right\}, \quad N \in \mathbb{N},$$

compare with (2.15) above. In [210, Thm. 4.9] it is shown

$$|\tilde{\Gamma}(N, d)| \leq e^2 N^{2 + \log_2 d}. \quad (10.3)$$

In [61] the authors state cardinality bounds (without proof) for slightly modified hyperbolic crosses in d dimensions.

What concerns the approximation in the uniform norm L_∞ in case $r > 1/2$ we refer to the recent paper Cobos, Kühn, Sickel [59]. Analyzing the formula (5.9) the authors obtained the asymptotic constant

$$\lim_{m \rightarrow \infty} \frac{m^{r-1/2} \lambda_m(\mathbf{W}_2^r, L_\infty)}{(\log m)^{(d-1)r}} = \frac{1}{\sqrt{2r-1}} \left[\frac{2^d}{(d-1)!} \right]^r.$$

Preasymptotic error bounds for isotropic periodic Sobolev spaces are given in [209], [208].

10.3 Approximation in the energy norm

Motivated by the aim to approximate the solution of a Poisson equation in the energy norm, i.e., in the norm of the isotropic Sobolev space H^1 , Bungartz and Griebel [44] investigated upper estimates of the quantities $\lambda_m(\mathbf{W}_2^r, H^1)$, where H_1 denotes the isotropic Sobolev space in L_2 , the so-called energy space. These studies have been continued in Griebel, Knappek [152, 153], Bungartz, Griebel [45], Griebel [151], Schwab, Süli, Todor [293], and Dinh Dũng, Ullrich [108]. A first result on upper bounds in this setting is due to Griebel

$$\lambda_m(\mathbf{W}_\infty^2([0, 1]^d), H^1) \leq C_d m^{-1} \quad , \quad m \in \mathbb{N}.$$

Note, that the usual log-term disappears. In other words, there is no bad d -dependence in the rate. For constructing an appropriate approximant we have to modify the standard hyperbolic cross according to the energy-norm. Instead of (2.15) we use projections onto the “energy hyperbolic cross”

$$Q_n^E := \bigcup_{r|\mathbf{s}|_1 - |\mathbf{s}|_\infty \leq n} \rho(\mathbf{s}).$$

Griebel [151] observed that in case $r = 2$ the constant $C_d = d^2 \cdot 0.97515^d$ suffices for large $m > m_d$. The precise range for m has not been given. The following theorem is stated in [108, Thm. 4.7(ii)].

Theorem 10.4. *Let $1 < r \leq 2$. Then one can precisely determine a threshold $\lambda = \lambda(r) > 1$ such that for $m > \lambda^d$, the correct upper and lower bounds explicit in the dimension d*

$$C_r d^{r-1} \left(\frac{1}{2^{1/(r-1)} - 1} \right)^d m^{-(r-1)} \leq \lambda_m(\tilde{\mathbf{W}}_2^r, H^1(\mathbb{T}^d)) \leq C'_r d^{r-1} \left(\frac{1}{2^{1/(r-1)} - 1} \right)^d m^{-(r-1)} \quad (10.4)$$

hold true with some explicit positive constants C_r, C'_r depending on r only.

Here $\tilde{\mathbf{W}}_2^r$ is the subspace of \mathbf{W}_2^r containing all functions f such that $\hat{f}(\mathbf{k}) \neq 0 \implies \prod_{i=1}^d k_i \neq 0$, and \mathbf{W}_2^r is defined as the set of function in $L_2(\mathbb{T}^d)$ for which the right hand side of (3.5) is finite. The upper bound in (10.4) also holds true for every $r > 1$. Note, that the terms depending on d in the both sides of (10.4) are the same and decay exponentially in d for $r < 2$. For $r = 2$, the relations (10.4) becomes

$$C_2 d m^{-1} \leq \lambda_m(\tilde{\mathbf{W}}_2^2, H^1(\mathbb{T}^d)) \leq C'_2 d m^{-1}.$$

For more results on this direction, see [108]. For a detailed comparison of the mentioned results we refer again to [210, Sect. 4.5].

10.4 ε -dimension and approximation in infinite dimensions

In computational mathematics, the so-called ε -dimension $n_\varepsilon = n_\varepsilon(\mathbf{F}, X)$ is used to quantify the computational complexity (in Information-Based Complexity (IBC) a similar object is termed *information complexity* or ε -cardinality). It is defined by

$$n_\varepsilon(\mathbf{F}, X) := \inf \left\{ m \in \mathbb{N} : \exists L_m : \sup_{f \in \mathbf{F}} \inf_{g \in L_m} \|f - g\|_X \leq \varepsilon \right\}, \quad (10.5)$$

where L_m denotes a linear subspace in X of dimension $\leq m$. This approximation characteristic is the inverse of $d_m(\mathbf{F}, X)$. In other words, the quantity $n_\varepsilon(\mathbf{F}, X)$ is the minimal number n_ε such

that the approximation of \mathbf{F} by a suitably chosen approximant n_ε -dimensional subspace L in X gives the approximation error $\leq \varepsilon$ (see [82, 83]). The quantity n_ε represents a modification of the information complexity (used in IBC) which is described by the minimal number $n(\varepsilon, d)$ of “linear” information (in case of Λ^{all}) needed to solve the corresponding d -variate linear approximation problem of the identity operator within accuracy ε . In contrast to (10.5) this quantity is defined as the inverse of the well-known Gelfand width, see [253, 4.1.4, 4.2] and (9.3) above. For further information on this topic we refer the interested reader to the books [253, 254, 255] and the references therein. The following theorem on ε -dimensions of the Sobolev class \mathbf{W}_2^r in the high-dimensional setting (see the previous subsection) has been proved in [56].

Theorem 10.5. *Let $r > 0$, $d \geq 2$. Then we have for every $\varepsilon \in (0, 1]$,*

$$n_\varepsilon(\mathbf{W}_2^r, L_2(\mathbb{T}^d)) \leq \frac{2^d}{(d-1)!} \varepsilon^{-1/r} (\ln \varepsilon^{-1/r} + d \ln 2)^{d-1}$$

and for every $\varepsilon \in (0, (2/3)^{rd})$

$$n_\varepsilon(\mathbf{W}_2^r, L_2(\mathbb{T}^d)) \geq \frac{2^d}{(d-1)!} \frac{\varepsilon^{-1/r} [\ln \varepsilon^{-1/r} - d \ln(3/2)]^d}{\ln \varepsilon^{-1/r} - d \ln(3/2) + d} - 1.$$

It is worth to emphasize that in high-dimensional approximation, the form of the upper and lower bounds for $n_\varepsilon(\mathbf{W}_2^r, L_2(\mathbb{T}^d))$ as in Theorem 10.5 is more natural and suitable than the form as used in the traditional form where the terms $\varepsilon^{-1/r} |\log \varepsilon|^{(s-1)/r}$ are a priori split from constants which are actually a function of dimension parameter d (and smoothness parameter r), and therefore, any high-dimensional estimate based on them leads to a rougher bound. The situation is analogous for Kolmogorov m -widths of classes of functions having a mixed smoothness if the terms $m^{-r} (\log m)^{r(d-1)}$ are a priori split from constants depending on the dimension d .

See also [108] for some similar results, and [56] for a general version of Theorem 10.5 as well an extension to non-periodic functions. From Theorem 10.5 we have

$$\lim_{\varepsilon \rightarrow 0} \frac{n_\varepsilon(\mathbf{W}_2^r, L_2(\mathbb{T}^d))}{\varepsilon^{-1/r} (\ln \varepsilon^{-1/r})^{d-1}} = \frac{2^d}{(d-1)!}.$$

The efficient approximation of a function of infinitely many variables is an important issue for a lot of problems in uncertainty quantification, computational finance and computational physics and is encountered for example in the numerical treatment of path integrals, stochastic processes, random fields and stochastic or parametric PDEs. While the problem of quadrature of functions in weighted Hilbert spaces with infinitely many variables has recently found a lot of interest in the information based complexity community, see e.g., [19, 66, 78, 79, 146, 147, 170, 171, 212, 246, 262, 263, 303, 402], there is much less literature on approximation. So far, the approximation of functions in weighted Hilbert spaces with infinitely many variables has been studied for a properly weighted L_2 -error norm in [400]. (See also [399, 403, 404], where however a norm in a special Hilbert space was employed such that the approximation problem indeed got easier than the integration problem.) It has been noticed in [257], one may have two options for obtaining tractability in infinite-dimensional approximation: either by using decaying weights or by using an increasing smoothness vector \mathbf{r} , see also Section 10.1 above.

In the recent decades, various approaches and methods have been proposed and studied for numerical solving of the parametric and stochastic elliptic PDE

$$-\operatorname{div}_{\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} u(\mathbf{x}, \mathbf{y})) = f(\mathbf{x}) \quad \mathbf{x} \in D \quad \mathbf{y} \in \Omega, \quad (10.6)$$

with homogeneous boundary conditions $u(\mathbf{x}, \mathbf{y}) = 0$, $\mathbf{x} \in \partial D$, $\mathbf{y} \in \Omega$, where $D \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^d$ and d may be very large or infinity. See to [60, 157, 292] for surveys and bibliography on different aspects in approximation and numerical methods for the problem (10.6).

The recent paper [104] has investigated the linear hyperbolic cross approximation in infinite dimensions of functions from spaces of mixed Sobolev type smoothness and mixed Sobolev-analytic-type smoothness in the infinite-variate case where specific summability properties of the smoothness are fulfilled. Such function spaces appear for example for the solution of the equation (10.6). The optimality of this approximation is studied in terms of the ε -dimension of their unit balls for which tight upper and lower bounds are given. These results then are applied to the linear approximation to the solution of the problem (10.6). The obtained upper and lower bounds of the approximation error as well as of the associated ε -dimension are completely independent of any parametric or stochastic dimension, and of the parameters which define the smoothness properties of the parametric or stochastic part of the solution. In the following, as an example let us briefly mention one of the results from [104] on the hyperbolic cross approximation in infinite dimensions in the norm of the space $\mathcal{G} := H^\beta(\mathbb{T}^m) \otimes L_2(\mathbb{T}^\infty)$ of functions from the space $\mathcal{H} := H^\alpha(\mathbb{T}^m) \otimes K^{\mathbf{r}}(\mathbb{T}^\infty)$, and its optimality in terms of ε -dimension $n_\varepsilon(\mathcal{U}, \mathcal{G})$. The space $H^\gamma(\mathbb{T}^m)$, $\gamma \geq 0$, equipped with a different equivalent norm can be identified with the isotropic Sobolev space $W_2^\gamma(\mathbb{T}^m)$. The space of anisotropic infinite-dimensional mixed smoothness space $K^{\mathbf{r}}(\mathbb{T}^\infty)$ with $\mathbf{r} = (r_1, r_2, \dots) \in \mathbb{R}^\infty$ and $0 < r_1 \leq r_2, \dots \leq r_j \dots$, is an infinite-variate generalization of the space $K^{\mathbf{r}}(\mathbb{T}^d)$ which equipped with a different equivalent norm can be identified with the space $\mathbf{W}_2^{\mathbf{r}}(\mathbb{T}^d)$ (see [104] for exact definitions of these spaces). Denote by \mathcal{U} the unit ball in \mathcal{H} .

Theorem 10.6. *If $(\alpha - \beta)/m < r_1$ and $\sum_{i=1}^\infty \frac{(3/2)^{1-mr_i/(\alpha-\beta)}}{mr_i/(\alpha-\beta)-1} < \infty$, we have for every $\varepsilon \in (0, 1]$,*

$$\lfloor \varepsilon^{-1/(\alpha-\beta)} \rfloor^m - 1 \leq n_\varepsilon(\mathcal{U}, \mathcal{G}) \leq C \varepsilon^{-m/(\alpha-\beta)}, \quad (10.7)$$

where C is a constant depending on $\alpha, \beta, m, \mathbf{r}$ only.

The upper bound in (10.7) is realized by a linear hyperbolic cross approximation in infinite dimensions corresponding to the infinite-variate anisotropic mixed smoothness of the spaces \mathcal{H} and \mathcal{G} . Depending on the regularity of the diffusions $\sigma(\mathbf{x}, \mathbf{y})$ and the right hand side $f(\mathbf{x})$ in the periodic equation (10.6) with $D = \mathbb{T}^m$ and $\Omega = \mathbb{T}^\infty$, we may assume the solution $u(\mathbf{x}, \mathbf{y})$ belonging to \mathcal{H} for some α and \mathbf{r} . Then we approximate $u(\mathbf{x}, \mathbf{y})$ in the “energy” norm of \mathcal{G} for some $0 \leq \beta < \alpha$. Based on the linear hyperbolic cross approximation in infinite dimension and (10.7), we can construct a linear method of rank $\leq n$ which gives the convergence rate of approximation to the solution of (10.6) as $n^{-(\alpha-\beta)/m}$. See [104, 102] for more results, on infinite dimensional approximation and applications in parametric and stochastic PDEs, in particular, on non-periodic and mixed versions of (10.6) and of Theorem 10.6.

11 Appendix

11.1 General notation

Let us start with introducing some notations. For $1 \leq p \leq \infty$ we shall denote by p' the duality exponent, that is, the number (or ∞) such that $1/p + 1/p' = 1$. For a vector $\mathbf{1} \leq \mathbf{p} \leq \infty$ we denote $\mathbf{p}' = (p'_1, \dots, p'_d)$ and $1/\mathbf{p} = (1/p_1, \dots, 1/p_d)$.

For the sake of brevity we shall write $\int f d\mu$ instead of $(2\pi)^{-d} \int_{\mathbb{T}^d} f(\mathbf{x}) d\mathbf{x}$, where $\mathbb{T}^d = [-\pi, \pi]^d$ and μ means the normalized Lebesgue measure on \mathbb{T}^d . For functions $f, g : \mathbb{T}^d \rightarrow \mathbb{C}$ we define the convolution

$$f * g(\mathbf{x}) := (2\pi)^{-d} \int_{\mathbb{T}^d} f(\mathbf{y}) g(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

In the case $\mathbf{p} = \mathbf{1}p$ we shall write the scalar p instead of the vector \mathbf{p} . Let further denote $L_q(\mathbb{T}^d)$, $0 < q \leq \infty$, the space of all measurable functions $f : \mathbb{T}^d \rightarrow \mathbb{C}$ satisfying

$$\|f\|_p := \left(\int_{\mathbb{T}^d} |f(\mathbf{x})|^p d\mu \right)^{1/p} < \infty$$

with the usual modification in case $p = \infty$. In case $1 \leq p \leq \infty$ the quantity $\|\cdot\|_p$ represents a norm. In case $0 < p < 1$ it is a quasi-norm. The space $C(\mathbb{T}^d)$ is often used as a replacement for $L_\infty(\mathbb{T}^d)$. It denotes the collection of all continuous periodic functions equipped with the L_∞ -topology.

11.2 Inequalities

We shall mention some well-known inequalities.

1.1. The Hölder inequality. Let $1 \leq p \leq \infty$, $f_1 \in L_p$, $f_2 \in L_{p'}$. Then $f_1 f_2 \in L_1$ and

$$\int |f_1 f_2| d\mu \leq \|f_1\|_p \|f_2\|_{p'}. \quad (11.1)$$

1.2. As a consequence of the relation (11.1) we obtain the Hölder inequality for a vector $\mathbf{1} \leq \mathbf{p} \leq \infty$

$$\int |f_1 f_2| d\mu \leq \|f_1\|_{\mathbf{p}} \|f_2\|_{\mathbf{p}'}$$

1.3. The Hölder inequality for several functions. Let $1 \leq p_i \leq \infty$, $i = 1, \dots, m$, $1/p_1 + \dots + 1/p_m = 1$, $f_i \in L_{p_i}$, $i = 1, \dots, m$. Then $f_1 \dots f_m \in L_1$ and

$$\int |f_1 \dots f_m| d\mu \leq \|f_1\|_{p_1} \dots \|f_m\|_{p_m}. \quad (11.2)$$

1.4. The monotonicity of L_p -norms. Let $1 \leq q \leq p \leq \infty$, then

$$\|f\|_q \leq \|f\|_p$$

and for $\mathbf{1} \leq \mathbf{q} \leq \mathbf{p} \leq \infty$,

$$\|f\|_{\mathbf{q}} \leq \|f\|_{\mathbf{p}}.$$

1.5. Let $1 \leq a < p < b \leq \infty$, $\theta = (1/p - 1/b)(1/a - 1/b)^{-1}$, then

$$\|f\|_p \leq \|f\|_a^\theta \|f\|_b^{1-\theta}.$$

1.6. From the inequality (11.1) we easily obtain the Hölder inequality for sums :

$$\sum_{i=1}^N |a_i b_i| \leq \left(\sum_{i=1}^N |a_i|^p \right)^{1/p} \left(\sum_{i=1}^N |b_i|^{p'} \right)^{1/p'}, \quad 1 \leq p \leq \infty.$$

We remark that in this inequality one can take $N = \infty$.

1.7. The Minkowski inequality. Let $1 \leq p \leq \infty$, $f \in L_p$, $i = 1, \dots, m$. Then

$$\left\| \sum_{i=1}^m f_i \right\|_p \leq \sum_{i=1}^m \|f_i\|_p.$$

1.8. It is possible to deduce the generalized Minkowski inequality from the Minkowski inequality. Let $1 \leq \mathbf{p} \leq \infty$, then

$$\left\| \int \varphi(\cdot, \mathbf{y}) d\mu(\mathbf{y}) \right\|_{\mathbf{p}} \leq \int \|\varphi(\cdot, \mathbf{y})\|_{\mathbf{p}} d\mu(\mathbf{y}).$$

1.9. Let $1 \leq q \leq p \leq \infty$, then

$$\left(\int \left(\int |f(\mathbf{x}, \mathbf{y})|^q d\mu(\mathbf{y}) \right)^{p/q} d\mu(\mathbf{x}) \right)^{1/p} \leq \left(\int \left(\int |f(\mathbf{x}, \mathbf{y})|^p d\mu(\mathbf{x}) \right)^{q/p} d\mu(\mathbf{y}) \right)^{1/q}.$$

1.10. The Young inequality. Let p, q and a be real numbers satisfying the conditions

$$1 \leq p \leq q \leq \infty, \quad 1 - 1/p + 1/q = 1/a.$$

Let $f \in L_p$ and $K \in L_a$ be 2π -periodic functions of a single variable. Let us consider the convolution of these functions

$$J(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} K(x-y) f(y) dy = K * f.$$

Then

$$\|J\|_q \leq \|K\|_a \|f\|_p.$$

1.11. The Young inequality for vector $\mathbf{p}, \mathbf{q}, \mathbf{a}$.

Let $1 \leq \mathbf{p} \leq \mathbf{q} \leq \infty$, $1 - 1/\mathbf{p} + 1/\mathbf{q} = 1/\mathbf{a}$

$$J(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) d\mu(\mathbf{y}) = K * f.$$

Then

$$\|J\|_{\mathbf{q}} \leq \|K\|_{\mathbf{a}} \|f\|_{\mathbf{p}}.$$

1.12. The Abel inequality. For nonnegative and non-increasing v_1, \dots, v_n we have

$$\left| \sum_{i=1}^n u_i v_i \right| \leq v_1 \max_k \left| \sum_{i=1}^k u_i \right|.$$

This inequality easily follows from the following formula

$$\sum_{i=1}^n u_i v_i = \sum_{\nu=1}^{n-1} (v_{\nu} - v_{\nu+1}) \sum_{i=1}^{\nu} u_i + v_n \sum_{i=1}^n u_i,$$

which is called the Abel transformation.

Along with spaces L_p we shall use spaces l_p , $1 \leq p \leq \infty$, of sequences $\mathbf{z} = \{z_k\}_{k=1}^\infty$ equipped with the norm

$$\|\mathbf{z}\|_p = \|\mathbf{z}\|_{l_p} := \left(\sum_{k=1}^{\infty} |z_k|^p \right)^{1/p}, \quad 1 \leq p < \infty,$$

$$\|\mathbf{z}\|_\infty := \|\mathbf{z}\|_{l_\infty} = \sup_k |z_k|.$$

The spaces l_p are Banach spaces.

11.3 Duality in L_p spaces

2.1. Let $f \in L_p$, $g \in L_{p'}$. We denote

$$\langle f, g \rangle := (2\pi)^{-d} \int_{\pi_d} f(\mathbf{x}) \overline{g(\mathbf{x})} d\mathbf{x} = \int f \bar{g} d\mu,$$

where \bar{z} is the complex conjugate number to a number z .

Theorem 11.1. *Let $1 \leq p \leq \infty$ and $f \in L_p$ then*

$$\|f\|_p = \sup_{g \in L_{p'}, \|g\|_{p'} \leq 1} |\langle f, g \rangle|.$$

Remark 11.2. *The statement analogous to Theorem 11.1 is valid for spaces l_p :*

$$\|\mathbf{z}\|_{l_p} = \sup_{\|\mathbf{w}\|_{l_{p'}} \leq 1} |(\mathbf{z}, \mathbf{w})|, \quad 1 \leq p \leq \infty.$$

2.2. Let F be a complex linear normed space and F^* be the conjugate (dual) space to F , that is elements of F^* are linear functionals φ defined on F with the norm

$$\|\varphi\| = \sup_{f \in F: \|f\| \leq 1} |\varphi(f)|.$$

Let $\Phi = \{\varphi_k\}_{k=1}^n$ be a set of functionals from F^* . Denote

$$F_\Phi = \{f \in F : \varphi_k(f) = 0, \quad k = 1, \dots, n\}.$$

Theorem 11.3. (The Nikol'skii duality theorem.) *Let $\Phi = \{\varphi_k\}_{k=1}^n$ be a fixed system of functionals from F^* . Then for any $\varphi \in F^*$*

$$\inf_{\{c_k\}_{k=1}^n} \left\| \varphi - \sum_{k=1}^n c_k \varphi_k \right\| = \sup_{f \in F_\Phi: \|f\| \leq 1} |\varphi(f)|.$$

Theorem 11.4. *Let $\varphi, \varphi_1, \dots, \varphi_n \in L_p$, $1 \leq p \leq \infty$, then*

$$\inf_{c_k; k=1, \dots, n} \left\| \varphi - \sum_{k=1}^n c_k \varphi_k \right\|_p = \sup_{\|g\|_{p'} \leq 1: (\varphi_k, g) = 0, k=1, \dots, n} |\langle \varphi, g \rangle|.$$

11.4 Fourier series of functions in L_p

3.1. For a function $f \in L_1(\mathbb{T}^d)$ we define Fourier coefficients

$$\hat{f}(\mathbf{k}) = (2\pi)^{-d} \int_{\mathbb{T}^d} f(\mathbf{x}) e^{-i(\mathbf{k}, \mathbf{x})} d\mathbf{x} = \langle f, e^{i(\mathbf{k}, \cdot)} \rangle, \quad \mathbf{k} \in \mathbb{Z}^d.$$

There are the well-known Parseval equality: for any $f \in L_2(\mathbb{T}^d)$

$$\|f\|_2 = \left(\sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|^2 \right)^{1/2},$$

and the Riesz-Fischer theorem: if $\sum_{\mathbf{k}} |c_{\mathbf{k}}|^2 < \infty$, then

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \in L_2 \quad \text{and} \quad \hat{f}(\mathbf{k}) = c_{\mathbf{k}}.$$

In the space L_p , $1 < p < \infty$, the following statement holds.

Theorem 11.5. (The Hausdorff-Young theorem.) *Let $1 < p \leq 2$, then for any $f \in L_p$,*

$$\left(\sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|^{p'} \right)^{1/p'} \leq \|f\|_p.$$

If a sequence $\{c_{\mathbf{k}}\}$ is such that $\sum_{\mathbf{k}} |c_{\mathbf{k}}|^p < \infty$, then there exists a function $f \in L_{p'}$ for which $\hat{f}(\mathbf{k}) = c_{\mathbf{k}}$ and

$$\|f\|_{p'} \leq \left(\sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|^p \right)^{1/p}.$$

This theorem can be derived from the following interpolation theorem, which is a special case of the general Riesz-Thorin theorem.

Denote the norm of an operator T acting from a Banach space E to a Banach space F by

$$\|T\|_{E \rightarrow F} = \sup_{\|f\|_E \leq 1} \|Tf\|_F.$$

Theorem 11.6. (The Riesz-Thorin theorem.) *Let E_q be either L_q or l_q and F_p be either L_p or l_p and for $1 \leq q_i, p_i \leq \infty$,*

$$\|T\|_{E_{q_i} \rightarrow F_{p_i}} \leq M_i, \quad i = 1, 2.$$

Then for all $0 < \theta < 1$

$$\|T\|_{E_q \rightarrow F_p} \leq M_1^\theta M_2^{1-\theta},$$

where

$$1/q = \theta/q_1 + (1 - \theta)/q_2, \quad 1/p = \theta/p_1 + (1 - \theta)/p_2.$$

3.2. Let $[y]$ be the integer part of the real number y , that is, the largest integer $[y]$ such that $[y] \leq y$. For a vector $\mathbf{s} = (s_1, \dots, s_d)$ with nonnegative integer coordinates we define the set $\rho(\mathbf{s})$ of vectors \mathbf{k} with integer coordinates:

$$\rho(\mathbf{s}) = \{\mathbf{k} = (k_1, \dots, k_d) : [2^{s_j-1}] \leq |k_j| < 2^{s_j}, \quad j = 1, \dots, d\}.$$

For $f \in L_1$ we denote

$$\delta_{\mathbf{s}}(f, \mathbf{x}) := \sum_{\mathbf{k} \in \rho(\mathbf{s})} \hat{f}(\mathbf{k}) e^{i(\mathbf{k}, \mathbf{x})}.$$

Theorem 11.7. (The Littlewood-Paley theorem.) Let $1 < p < \infty$. There exist positive numbers $C_1(d, p)$ and $C_2(d, p)$, which depend on d and p , such that for each function $f \in L_p$,

$$C_1(d, p) \|f\|_p \leq \left\| \left(\sum_{\mathbf{s}} |\delta_{\mathbf{s}}(f, \mathbf{x})|^2 \right)^{1/2} \right\|_p \leq C_2(d, p) \|f\|_p.$$

Corollary 11.8. Let G be a finite set of vectors \mathbf{s} and let the operator S_G map a function $f \in L_p$, $p > 1$, to a function

$$S_G(f) = \sum_{\mathbf{s} \in G} \delta_{\mathbf{s}}(f).$$

Then

$$\|S_G\|_{L_p \rightarrow L_p} \leq C(d, p), \quad 1 < p < \infty.$$

For the sake of brevity we shall write $\|T\|_{L_q \rightarrow L_p} = \|T\|_{q \rightarrow p}$.

Corollary 11.9. Let $p^* = \min\{p, 2\}$; then for $f \in L_p$ we have

$$\|f\|_p \leq C(d, p) \left(\sum_{\mathbf{s}} \|\delta_{\mathbf{s}}(f, \mathbf{x})\|_p^{p^*} \right)^{1/p^*}, \quad 1 < p < \infty.$$

Theorem 11.10. (The Marcinkiewicz multiplier theorem.) Suppose that $\lambda_0, \lambda_1, \dots$ are Marcinkiewicz multipliers, that is, they satisfy the conditions

$$|\lambda_n| \leq M, \quad n = 0, \pm 1, \dots; \quad \sum_{l=\pm 2^\nu}^{\pm(2^{\nu+1}-1)} |\lambda_l - \lambda_{l+1}| \leq M, \quad \nu = 0, 1, \dots,$$

where M is a number.

Then the operator Λ which maps a function f to a function

$$\sum_k \lambda_k \hat{f}(k) e^{ikx},$$

is bounded as operator from L_p to L_p for $1 < p < \infty$.

Theorem 11.11. (The Hardy-Littlewood-Sobolev inequality.) Let $1 < q < p < \infty$,

$$\mu = 1 - 1/q + 1/p, \quad \|f\|_{L_q(\mathbb{R})} = \left(\int_{-\infty}^{\infty} |f(x)|^q dx \right)^{1/q} < \infty, \quad Jf(x) := \int_{-\infty}^{\infty} f(y) |x-y|^{-\mu} dy.$$

Then the inequality

$$\|Jf\|_{L_p(\mathbb{R})} \leq C(q, p) \|f\|_{L_q(\mathbb{R})}$$

holds.

Corollary 11.12. Let $1 < q < p < \infty$, $\beta = 1/q - 1/p$. Then the operator A_β which maps a function $f \in L_q$ to a function

$$\sum_{\mathbf{k}} \hat{f}(\mathbf{k}) \left(\prod_{j=1}^d \max\{1, |k_j|\} \right)^{-\beta} e^{i(\mathbf{k}, \mathbf{x})}$$

is bounded as operator from L_q to L_p .

References

- [1] C. Aistleitner, Covering numbers, dyadic chaining and discrepancy, *J. Complexity* **27** (2011) 531–540.
- [2] G.A. Akishev, Approximation of function classes in spaces with mixed norm, *Mat. Sb.*, **197** (2006), 17–40.
- [3] G.A. Akishev, On the exact estimations of the best M -terms approximation of the Besov class, *Sib. Elektron. Mat. Izv.*, **7** (2010), 255–274.
- [4] T.I. Amanov, Spaces of differentiable functions with dominating mixed derivatives. Nauka Kaz. SSR, Alma-Ata, 1976.
- [5] A.V. Andrianov and V.N. Temlyakov, On two methods of generalization of properties of univariate function systems to their tensor product, *Trudy MIAN*, **219** (1997), 32–43; English transl. in *Proceedings of the Steklov Inst. Math.*, **219** (1997), 25–35.
- [6] K.I. Babenko, On the approximation of periodic functions of several variables by trigonometric polynomials, *Dokl. Akad. Nauk USSR*, **132** (1960), 247–250; English transl. in *Soviet Math. Dokl.*, **1** (1960).
- [7] K.I. Babenko, On the approximation of a certain class of periodic functions of several variables by trigonometric polynomials, *Dokl. Akad. Nauk USSR*, **132** (1960), 982–985; English transl. in *Soviet Math. Dokl.*, **1** (1960).
- [8] K.I. Babenko, Some problems in approximation theory and numerical analysis, *Russian Math. Surveys*, **40** (1985), 1–30.
- [9] M. Bachmayr, A. Cohen, R. DeVore, G. Migliorati, Sparse polynomial approximation of parametric elliptic PDEs. Part II: lognormal coefficients, *ESAIM Math. Model. Numer. Anal.*, DOI=10.1051/m2an/2016051 (to appear), *Arxiv e-print* arxiv:arXiv:1509.07050.
- [10] M. Bachmayr, W. Dahmen, R. DeVore, and L. Grasedyck, Approximation of high-dimensional rank one tensors, *Constr. Approx.*, **39**(2):385–395, 2014.
- [11] E. A. Bailov, M. B. Sikhov, N. Temirgaliev, General algorithm for the numerical integration of functions of several variables, *Zh. Vychisl. Mat. Mat. Fiz.*, **54** (2014), 1059–1077.
- [12] R.C. Baker, On irregularities of distribution. II, *J. Lond. Math. Soc. Ser. II* **59**(1) (1999) 50–64.
- [13] Sh. A. Balgimbaeva, Nonlinear approximation of function spaces of mixed smoothness, *Sibirsk. Mat. Zh.* **56** (2015), 322–337; English translation in *Sib. Math. J.* **56** (2015), 262–274.
- [14] Sh. A. Balgimbaeva and T.I. Smirnov, Estimates of the Fourier widths of classes of periodic functions with mixed modulus of smoothness, *Tr. Inst. Math. and Mekh. UrO RAN*, **21** (2015), 78–94.
- [15] N.S. Bakhvalov, Imbedding theorems for classes of functions with several bounded derivatives, *Vestnik Moskov. Univ. Ser. Mat. Mekh.* no.3, 7–16, 1963.

- [16] N.S. Bakhvalov, On the approximate calculation of multiple integrals, *Vestnik Moskov. Univ. Ser. Mat. Mekh. Astr. Fiz. Khim.*, 1959, issue 4, 3–18 (see also *J. Complexity*, **31**, 502–516, 2015).
- [17] N.S. Bakhvalov, Optimal convergence bounds for quadrature processes and integration methods of Monte Carlo type for classes of functions, *Zh. Vychisl. Mat. i Mat. Fiz. Suppl.* **4** (1963), issue 4, 5–63.
- [18] N.S. Bakhvalov, Lower estimates of asymptotic characteristics of classes of functions with dominant mixed derivative, *Matem. Zametki*, **12** (1972), 655–664; English transl. in, *Math. Notes* **12** (1972).
- [19] J. Baldeaux and M. Gnewuch, Optimal randomized multilevel algorithms for infinite-dimensional integration on function spaces with ANOVA-type decomposition, *SIAM Journ. Num. Anal.*, **52**(3), (2014), 1128–1155.
- [20] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids. Multivariate polynomial interpolation, *Adv. Comput. Math.* **12**(4), pp. 273–288, 2000.
- [21] R.F. Bass, Probability estimates for multi-parameter Brownian processes, *Ann. Probab.*, **16** (1988), 251–264.
- [22] D. B. Bazarkhanov, Characterizations of the Nikol’skii-Besov and Lizorkin-Triebel Function Spaces of Mixed Smoothness, *Tr. Mat. Inst. Steklov*, **243** (2003), 53–65.
- [23] D. B. Bazarkhanov, Estimates of the Fourier Widths of Classes of Nikolskii-Besov and Lizorkin-Triebel Types of Periodic Functions of Several Variables, *Mat. Zametki*, **87** (2010), 305–308.
- [24] D. B. Bazarkhanov, Wavelet approximation and Fourier widths of classes of periodic functions of several variables. I, *Tr. Mat. Inst. Steklov*, **269** (2010), 8–30.
- [25] D. B. Bazarkhanov, Nonlinear approximations of classes of periodic functions of many variables *Tr. Mat. Inst. Steklov*, **284** (2014), 8–37.
- [26] D. Bazarkhanov and V.N. Temlyakov, Nonlinear tensor product approximation of functions, *J. Complexity*, **31**(6), 867–884, 2015. arXiv e-print:1409.1403v1 [stat.ML].
- [27] J. Beck and W. Chen, *Irregularities of distribution*, Cambridge University Press, Cambridge, 1987.
- [28] R. Bellmann, *Dynamic Programming*, Princeton University Press, Princeton, 1957.
- [29] E.S. Belinskii, Approximation by a “floating” system of exponentials on classes of periodic functions with bounded mixed derivative, *Research on the theory of functions of many real variables*, Yaroslavl’ State University, 1988, 16–33 (in Russian).
- [30] E.S. Belinskii, Approximation of functions of several variables by trigonometric polynomials with given number of harmonics, and estimates of ϵ -entropy, *Analysis Mathematica*, **15** (1989), 67–74.

- [31] E.S. Belinskii, Asymptotic characteristics of classes of functions with conditions on the mixed derivative (mixed difference) (Russian), Studies in the theory of functions of several real variables, in “Proceedings of Yaroslav University” (Y. Brudnyi, Ed.), pp. 22–37, Yaroslav, 1990.
- [32] E.S. Belinskii, Estimates for the Kolmogorov widths of the classes with conditions on mixed difference in the uniform metric, *Mat. Zametki* **50**(5), pp. 147–149, 1991 (Russian).
- [33] E.S. Belinskii, Decomposition theorems and approximation by a “floating” system of exponentials, Transactions of the American Mathematical Society, **350** (1998), 43–53.
- [34] E.S. Belinskii, Estimates of entropy numbers and Gaussian measures for classes of functions with bounded mixed derivative, *J. Approx. Theory*, **93** (1998), 114–127.
- [35] O.V. Besov, V.P. Il’in and S.M. Nikol’skii, Integral representations of functions and imbedding theorems, Vol. 1. Halsted Press, 1978.
- [36] D. Bilyk and M. Lacey, On the Small Ball Inequality in three dimensions, *Duke Math. J.*, **143** (2008), 81–115.
- [37] D. Bilyk, M. Lacey and A. Vagharshakyan, On the Small Ball Inequality in all dimensions, *J. Func. Analysis*, **254** (2008), 2470–2502.
- [38] P. Binev, A. Cohen, W. Dahmen, R. DeVore, and V.N. Temlyakov, Universal algorithms for learning theory. Part I: piecewise constant functions, Journal of Machine Learning Theory (JMLT), **6** (2005), 1297–1321.
- [39] C. de Boor, G.J. Fix, Spline approximation by quasi-interpolants, *J. Approx. Theory* **8**(1973), 19–45.
- [40] A. I. Borevich and I. R. Shafarevich. *Number theory*. Translated from the Russian by Newcomb Greenleaf. Pure and Applied Mathematics, Vol. 20. Academic Press, New York-London, 1966.
- [41] J. Bourgain and V.D. Milman, New volume ratio properties for convex symmetric bodies in \mathbb{R}^n , *Invent. Math.*, **88** (1987), 319–340.
- [42] S. Bugrov, Approximation of a class of functions with a dominant mixed derivative, *Mat. Sb.* **64**(106) (1964), 410–418.
- [43] H.-J. Bungartz, Finite elements of higher order on sparse grids, Habilitationsschrift, Institut für Informatik, TU München, and Shaker Verlag, Aachen, 1998.
- [44] H.-J. Bungartz and M. Griebel, A note on the complexity of solving Poisson’s equation for spaces of bounded mixed derivatives, *J. Complexity* **15** (1999), 167–199.
- [45] H.-J. Bungartz and M. Griebel, Sparse grids, *Acta Numerica* **13** (2004), 147–269.
- [46] V.A. Bykovskii, On the correct order of the error of optimal cubature formulas in spaces with dominant derivative, and on quadratic deviations of grids info Preprint, Computing Center Far-Eastern Scientific Center, Akad. Sci. USSR, Vladivostok, 1985.

- [47] G. Byrenheid, D. Dũng, W. Sickel, and T. Ullrich. Sampling on energy-norm based sparse grids for the optimal recovery of Sobolev type functions in H^γ , *J. Approx. Theory* **207**, pp. 207–231 (2016), *ArXiv e-prints*, 2014. arXiv:1408.3498 [math.NA].
- [48] G. Byrenheid and T. Ullrich, Optimal sampling recovery of mixed order Sobolev embeddings via discrete Littlewood-Paley type characterizations, arxiv e-print 2016, arXiv:1603.04809.
- [49] G. Byrenheid and T. Ullrich, The Faber-Schauder basis in spaces with bounded mixed derivative and nonlinear approximation, Preprint 2015.
- [50] G. Byrenheid, L. Kämmerer, T. Ullrich, T. Volkmer, Tight error bounds for rank-1 lattice sampling in spaces of hybrid mixed smoothness, arXiv:1510.08336v2, 2015 [math.NA].
- [51] B. Carl, Entropy numbers, s -numbers, and eigenvalue problem, *J. Func. Analysis*, **41** (1981), 290–306.
- [52] B. Carl and A. Pajor, Gel’fand numbers of operators with values in a Hilbert space, *Invent. Math.*, **94**(3):479–504, 1988.
- [53] B. Carl and I. Stephani, Entropy, Compactness and the Approximation of Operators, Cambridge 1990.
- [54] W.W.L. Chen, On irregularities of distribution, *Mathematika*, **27** (1980), 153–170.
- [55] W. W. L. Chen, M. M. Skriganov, Explicit constructions in the classical mean squares problem in irregularities of point distribution, *J. Reine Angew. Math.* **545** (2002), 67–95.
- [56] A. Chernov and D. Dũng, New explicit-in-dimension estimates for the cardinality of high-dimensional hyperbolic crosses and approximation of functions having mixed smoothness, *J. Complexity* **1**(2016), 92–121, arXiv:1309.5170v3 [math.NA]. .
- [57] C.K. Chui, An Introduction to Wavelets, Academic Press, New York, 1992.
- [58] C.K. Chui, H. Diamond, A natural formulation of quasi-interpolation by multivariate splines, *Proc. Amer. Math. Soc.* **99**(1987), 643–646.
- [59] F. Cobos, T. Kühn, W. Sickel, Optimal approximation of Sobolev functions in the sup-norm, *J. Funct. Anal.* **270**(11), pp. 4196–4212, 2016.
- [60] A. Cohen and R. DeVore, Approximation of high-dimensional parametric PDEs, *Acta Numerica* **24** (2015), 1–159.
- [61] A. Cohen, R. DeVore, S. Foucart, H. Rauhut, Recovery of functions of many variables via compressive sensing, in: *Proc. SampTA 2011*, Singapore, 2011.
- [62] A. Cohen, R. DeVore, C. Schwab, Analytic regularity and polynomial approximation of parametric and stochastic elliptic pdes, *Analysis and Applications* **9**(1) (2011), 11–47 .
- [63] P. Combettes and D. Dũng, Kolmogorov n -widths of function classes induced by a non-degenerate differential operator: A convex duality approach, *Set-Valued and Variational Analysis* **24**(1), 83–99 (2015), DOI 10.1007/s11228-015-0338-3.

- [64] J.G. van der Corput, Verteilungsfunktionen.I, Proc. Kon. Ned. Akad. v. Wetensch., **38** (1935), 813–821.
- [65] J.G. van der Corput, Verteilungsfunktionen.II, Proc. Kon. Ned. Akad. v. Wetensch., **38** (1935), 1058–1066.
- [66] J. Creutzig, S. Dereich, T. Müller-Gronbach, and K. Ritter, Infinite-dimensional quadrature and approximation of distributions, *Foundations of Computational Mathematics*, **9**, (2009), 391–429.
- [67] H. Davenport, Note on irregularities of distribution, *Mathematika*, **3** (1956), 131–135.
- [68] F.-J. Delvos and W. Schempp, *Boolean methods in interpolation and approximation*, Longman Scientific & Technical, Harlow, 1989.
- [69] R.A. DeVore, Nonlinear approximation *Acta Numerica* **7** (1998), 51–150.
- [70] R. DeVore, R. Howard and C. Micchelli, Optimal nonlinear approximation, *Manuscripta Mathematica* **63**(1989), 469–478.
- [71] R.A. DeVore, S.V. Konyagin, and V.N. Temlyakov, Hyperbolic Wavelet Approximation, *Constructive Approximation*, **14** (1998), 1–26.
- [72] R.A. DeVore, G. Kyriazis, D. Leviatan and V. Tikhomirov, Wavelet compression and nonlinear n -widths, *Adv. Comp. Math.* **1**(1993), 194–214.
- [73] R.A. DeVore and G.G. Lorentz, *Constructive approximation*, Springer, Berlin, 1993.
- [74] R.A. DeVore, P. P. Petrushev, and V.N. Temlyakov, Multivariate trigonometric approximation with frequencies from the hyperbolic cross, *Matem. Zametki*, **56** (1994), 36–63.
- [75] R.A. DeVore and V.N. Temlyakov, Nonlinear approximation by trigonometric sums, *J. Fourier Analysis and Applications*, **2** (1995), 29–48.
- [76] R.A. DeVore and V.N. Temlyakov, Some remarks on Greedy Algorithms, *Advances in Computational Mathematics*, **5** (1996), 173–187.
- [77] J. Dick. Explicit constructions of quasi-Monte Carlo rules for the numerical integration of high-dimensional periodic functions. *SIAM Journ. Num. Anal.*, **45**:2141–2176, 2007.
- [78] J. Dick and M. Gnewuch, Optimal randomized changing dimension algorithms for infinite-dimensional integration on function spaces with ANOVA-type decomposition, *Journal of Approximation Theory*, **184**, (2014), 111–145.
- [79] J. Dick and M. Gnewuch, Infinite-dimensional integration in weighted Hilbert spaces: Anchored decompositions, optimal deterministic algorithms, and higher-order convergence, *Foundations of Computational Mathematics*, **14**(5), (2014), 1027–1077.
- [80] J. Dick, F.Y. Kuo and I.H. Sloan, High-dimensional integration: The quasi-Monte Carlo way, *Acta Numerica* **22**, 133–288 (2013).
- [81] S.J. Dilworth, D. Kutzarova, V.N. Temlyakov, Convergence of some greedy algorithms in Banach spaces, *J. Fourier Analysis and Applications* **8** (2002), 489–505.

- [82] Dinh Dũng, Some approximative characteristics of the classes of smooth functions of several variables in the metric of L_2 , *Uspekhi Mat. Nauk*, **34**(1979), 189–190.
- [83] Dinh Dũng, Mean ε -dimension of the functional class $B_{G,p}$, *Mat. Zametki*, **28**(1980), 727–736.
- [84] Dinh Dũng, The number of integer points in a certain set and approximation of functions of several variables, *MZ* 36 (1984), 479–491; English transl. in *MN* 36 (1984).
- [85] Dinh Dũng, Approximation of classes of smooth functions of several variables, *Trudy Sem. Petrovsk.* 10 (1984), 207–226; English transl. in *J. Soviet Math.* 33 (1986), no. 6.
- [86] Dinh Dũng, Approximation of classes of functions on the torus prescribed by a mixed modulus of continuity,, in *Constructive Theory of Functions* (Proc. Internat. Conf., Varna, 1984), Publ. House Bulgarian Acad. Sci., Sofia, 1984, pp. 43–48. (Russian)
- [87] Dinh Dũng, Approximation of multivariate functions by means of harmonic analysis, *Hab. Dissertation* (1985), Moscow, MGU.
- [88] Dinh Dũng, Approximation by trigonometric polynomials of functions of several variables on the torus, *Mat Sb* 131(173)(1986), 251–271; English transl. in *Mat. Sb.* 59(1988).
- [89] Dinh Dũng, Best linear methods of approximation of classes of periodic functions of several variables, *Mat. Zametki* **41**(1987), 646–653.
- [90] Dinh Dũng, On recovery and one-sided approximation of periodic functions of several variables, *Dokl. Akad. SSSR* **313**(1990), 787–790.
- [91] Dinh Dũng, On optimal recovery of multivariate periodic functions, In: *Harmonic Analysis* (Conference Proceedings, Ed. S. Igary), Springer-Verlag 1991, Tokyo-Berlin, pp. 96–105.
- [92] Dinh Dũng, Best multivariate approximations by trigonometric polynomials with frequencies from hyperbolic crosses, *J. of Approx. Theory* **91**(1997), 205–225.
- [93] Dinh Dũng, Moduli of smoothness for hyperbolic approximation of multivariate functions, *East J. on Approx.* **3**(1997), 101–116.
- [94] Dinh Dũng, On nonlinear n -widths and n -term approximation, *Vietnam J. Math.* **26**(2)(1998), 165–176.
- [95] Dinh Dũng, Continuous algorithms in n -term approximation and non-linear widths, *Journ. Appr. Theor.* **102** (2000), 217–242.
- [96] Dinh Dũng, Non-linear approximations using sets of finite cardinality or finite pseudo-dimension, *J. Complexity* **17**(2001), 467–492.
- [97] Dinh Dũng, Asymptotic orders of optimal non-linear approximations, *East Journ. Appr.* **7**(2001), 55–76.
- [98] Dinh Dũng, Non-linear n -term approximations of smooth functions using wavelet decompositions, In: *Proc. of the Third Asian Mathematical Conference* (Eds. T. Sunada, P.W. Sy, Y. Lo), World Scientific, Singapore 2002, pp. 153–174.

- [99] Dinh Dũng, B-spline quasi-interpolant representations and sampling recovery of functions with mixed smoothness, *Journal of Complexity* **27**(2011), 541–467.
- [100] Dinh Dũng, Sampling and cubature on sparse grids based on a B-spline quasi-interpolation, *Found. Comp. Math.* **16** (2016), 1193–1240.
- [101] Dinh Dũng, High-dimensional periodic sampling on Smolyak grids based on B-spline quasi-interpolation, arXiv:1502.01447v2 [math.NA] (2015).
- [102] Dinh Dũng, Linear collective collocation and Galerkin approximations for parametric and stochastic elliptic PDEs, arxiv e-print 2015, arXiv:1511.03377.
- [103] Dinh Dũng, B-spline quasi-interpolation sampling representation and sampling recovery in Sobolev spaces of mixed smoothness, arxiv e-print 2016, arXiv:1603.01937.
- [104] Dinh Dũng and M. Griebel, Hyperbolic cross approximation in infinite dimensions, *J. Complexity* **33**, 55–88 (2016), DOI 10.1016/j.jco.2015.09.006.
- [105] Dinh Dũng and C. Micchelli, Multivariate approximation by translates of the Korobov function on Smolyak grids, *J. of Complexity* **29** (2013), pp. 424–437.
- [106] Dinh Dũng and C. Micchelli, Corrigendum to “Multivariate approximation by translates of the Korobov function on Smolyak grids” [*J. Complexity* 29 (2013) 424–437], *Journal of Complexity* **35** (2016), pp. 124–125.
- [107] Dinh Dũng and V.M. Tikhomirov, On inequalities for derivatives in L_2 -metric, *Vetsnik Moskov. Univ., Ser.I Mat. Mech.* No5(1979), 7–11.
- [108] Dinh Dũng and T. Ullrich, N -Widths and ε -dimensions for high-dimensional approximations, *Foundations of Computational Mathematics*, 13(2013), 965–1003.
- [109] Dinh Dũng and T. Ullrich, Lower bounds for the integration error for multivariate functions with mixed smoothness and optimal Fibonacci cubature for functions on the square, *Math. Nachr.* **288**(2015), 743–762.
- [110] Dinh Dũng and Vu Quoc Thanh, On non-linear n -widths, *Proc. Amer. Math. Soc.* **124**(1996), 2757–2763.
- [111] B. Doerr, A lower bound for the discrepancy of a random point set, *J. Complexity* **30** (2014), 16–20.
- [112] M. Döhler, S. Kunis and D. Potts, Nonequispaced hyperbolic cross fast Fourier transform. *SIAM Journ. Num. Anal.* **47**(6), 4415–4428, 2010.
- [113] M. Drmota, R.F. Tichy, Sequences, Discrepancies and Applications, in: *Lecture Notes in Mathematics*, Vol. 1651, Springer, Berlin, 1997.
- [114] V.V. Dubinin, Cubature formulas for classes of functions with bounded mixed difference, *Matem. Sbornik*, **183** (1992), issue 7; English transl. in *Math. USSR Sbornik* **76** (1993), 283–292.
- [115] V.V. Dubinin, Cubature formulae for Besov classes, *Izvestiya Math* **61**(2) (1997), 259–83.

- [116] T. Dunker, W. Linde, T. Kühn and M. Lifshits, Metric Entropy of Integration Operators and Small Ball Probabilities for the Brownian Sheet, *J. Approx. Theory*, **101** (1999), 63–77.
- [117] D.E. Edmunds and Yu. Netrusov, Entropy numbers of embeddings of Sobolev spaces in Zygmund spaces, *Studia Mathematica* **128**(1), pp. 71–102, 1998.
- [118] D.E. Edmunds and H. Triebel, Entropy numbers and approximation numbers in function spaces, II, *Proc. London Math. Soc.* (3) **64**, 153–169, 1992.
- [119] D.E. Edmunds and H. Triebel, *Function Spaces, Entropy Numbers and Differential Operators*, Cambridge 1996.
- [120] G. Faber. Über stetige Funktionen. *Math. Ann.*, 66:81–94, 1909.
- [121] H. Faure, Discrèpances de suites associées á un système de numéraction (en dimension s), *Acta Arith.*, **41** (1982), 337–351.
- [122] M. Fenn, S. Kunis, and D. Potts, Fast evaluation of trigonometric polynomials from hyperbolic crosses. *Numer. Algorithms* **41**(4), 339–352, 2006.
- [123] S. Foucart, Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants. In *Approximation Theory XIII: San Antonio 2010*, ed. by M. Neamtu, L. Schumaker. Springer Proceedings in Mathematics, vol. 13 (Springer, New York, 2012), pp. 65–77.
- [124] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich, The Gelfand widths of ℓ_p -balls for $0 < p \leq 1$. *J. Complexity*, **26**:629–640, 2010.
- [125] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.
- [126] J. Franke, On the spaces $F_{p,q}^s$ of Triebel-Lizorkin type: pointwise multipliers and spaces on domains, *Math. Nachr.* **125** (1986), 29–68.
- [127] M. Frazier and B. Jawerth, A discrete transform and decompositions of distribution spaces. *Journ. of Funct. Anal.*, **93**:34–170, 1990.
- [128] K.K. Frolov, Upper bounds on the error of quadrature formulas on classes of functions, *Dokl. Akad. Nauk SSSR* **231** (1976), 818–821; English transl. in *Soviet Math. Dokl.*, **17** (1976).
- [129] K.K. Frolov, Quadrature formulas on classes of functions, PhD dissertation, Vychisl. Tsentr Acad. Nauk SSSR, Moscow, 1979, in Russian.
- [130] K.K. Frolov, An upper estimate of the discrepancy in the L_p -metric, $2 \leq p < \infty$, *Dokl. Akad. Nauk SSSR*, **252**, (1980), 805–807; English transl. in, *Soviet Math. Dokl.*, **21** (1980).
- [131] E.M. Galeev, Approximation of classes of functions with several bounded derivatives by Fourier sums, *Mat. Zametki* **23**(1978), 197–212; English transl. in *MN* **23** (1978).
- [132] E.M. Galeev, The Kolmogorov diameter of the intersection of classes of periodic functions and of finite-dimensional sets *Mat. Zametki*, **29**(1981), 749–760.

- [133] E.M. Galeev, Order estimates of derivatives of the multidimensional periodic Dirichlet α -kernel in a mixed norm, *Mat. Sb.* **117(159)** (1982), 32–43; English transl. in *Mat. Sb.* **45** (1983).
- [134] E.M. Galeev, Kolmogorov widths of certain classes of periodic functions of several variables, in *Constructive Theory of Functions (Proc. Internat. Conf., Varna, 1984)*, Publ. House Bulgarian Acad. Sci., Sofia, 1984, pp. 27–32. (Russian).
- [135] E.M. Galeev, Kolmogorov widths of the classes $W_p^{\bar{\alpha}}$ and $H_p^{\bar{\alpha}}$ of periodic functions of several variables in the space \tilde{L}_q , *Izv* 49 (1985), 916–934; English transl. in *MIzv* 27(1986).
- [136] E.M. Galeev Orders of orthogonal projection widths of classes of periodic functions of one and several variables, *Matem. Zametki* **43** (1988), 197–211; English transl. in *Math. Notes* **43** (1988).
- [137] E.M. Galeev, On linear widths of classes of periodic functions of several variables *Vestnik MGU, Ser.1 Mat.-Mekh.*, No 4(1987), 13–16.
- [138] E.M. Galeev, Kolmogorov widths of classes of periodic functions of one and several variables *Izv. Akad. Nauk SSSR Ser. Mat.*, 54:2 (1990), 418–430.
- [139] E.M. Galeev, Approximation of classes of periodic functions of several variables by nuclear operators, *Mat. Zametki*, 47(1990), 32–41.
- [140] E.M. Galeev, Linear widths of Hölder-Nikol'skii classes of periodic functions of several variables *Mat. Zametki*, 59(1996), 189–199.
- [141] E.M. Galeev, Widths of the Besov classes $B_{p,\theta}^r(\mathbb{T}^d)$, *Mat. Zametki* 69(2001), 656–665.
- [142] A. Y. Garnaev and E. D. Gluskin, On widths of the Euclidean ball, *Sov. Math., Dokl.*, **30** (1984), 200–204.
- [143] T. Gerstner and M. Griebel, Numerical integration using sparse grids, *Numer. Algorithms* **18**(3-4) (1998), 209–232.
- [144] E.D. Gluskin, Norms of random matrices and widths of finite-dimensional sets, *Mat. Sb.* **120(162)** (1983), 180–189; English transl. in *Math. Sb.* **48** (1984).
- [145] E.D. Gluskin, Extremal properties of orthogonal parallelepipeds and their application to the geometry of Banach spaces, *Math USSR Sbornik* **64** (1989), 85–96.
- [146] M. Gnewuch, Infinite-dimensional integration on weighted Hilbert spaces, *Mathematics of Computation*, 81, (2012), 2175–2205.
- [147] M. Gnewuch, S. Mayer, and K. Ritter, On weighted Hilbert spaces and integration of functions of infinitely many variables, *Journal of Complexity*, 30(2), (2014), 29–47.
- [148] M. Gnewuch and H. Woźniakowski, Quasi-polynomial tractability, *J. Complexity* **27** (3-4) (2011), 312–330.
- [149] T. Goda, K. Suzuki, and T. Yoshiki, Optimal order quasi-Monte Carlo integration in weighted Sobolev spaces of arbitrary smoothness. *Arxiv e-print* arXiv:1508.06373, 2015.
- [150] T. Goda, K. Suzuki, and T. Yoshiki, An explicit construction of optimal order quasi-Monte Carlo rules for smooth integrands. *Arxiv e-print* arXiv:1601.06501, 2016.

- [151] M. Griebel, Sparse grids and related approximation schemes for higher dimensional problems. In: Proceedings of Foundations of Computational Mathematics, Santander 2005, pp. 106–161. London Mathematical Society Lecture Notes Series, 331, Cambridge University Press, Cambridge (2006)
- [152] M. Griebel and S. Knapek (2000), ‘Optimized tensor-product approximation spaces’, *Constr. Approx.* **16**, 525–540.
- [153] M. Griebel and S. Knapek, Optimized general sparse grid approximation spaces for operator equations. *Math. Comp.*, **78**(268):2223–2257, 2009.
- [154] M. Griebel, M. Schneider and C. Zenger, A combination technique for the solution of sparse grid problems, in *Iterative Methods in Linear Algebra* (P. de Groen and R. Beauwens, eds), Elsevier, Amsterdam, pp. 263–281, 1992.
- [155] P. M. Gruber and C. G. Lekkerkerker. *Geometry of numbers*, volume 37 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1987.
- [156] O. Guédon and A. E. Litvak, Euclidean projections of a p -convex body. In *Geometric Aspects of Functional Analysis*, volume 1745 of *Lecture Notes in Math.* Springer, Berlin, 2000.
- [157] M. Gunzburger, C. Webster and G. Zang, Stochastic finite element methods for partial differential equations with random input data, *Acta Numerica* **23**, 521–650 (2014).
- [158] L. Györfy, M. Kohler, A. Krzyzak, and H. Walk, *A distribution-free theory of nonparametric regression*, Springer, Berlin, 2002.
- [159] G. Halász, On Roth’s method in the theory of irregularities of points distributions, *Recent Progress in Analytic Number Theory*, **2** (1981), 79–94.
- [160] K. Hallatschek, Fouriertransformation auf dünnen Gittern mit hierarchischen Basen, *Numer. Math.* **63**, pp. 83–97, 1992.
- [161] J.H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* **2** (1960) 84–90, 196.
- [162] J.H. Halton and S.K. Zaremba, The extreme and L_2 discrepancies of some plane sets, *Monats. für Math.*, **73** (1969), 316–328.
- [163] J.M. Hammersley, Monte Carlo methods for solving multivariable problems. *Ann. New York Acad. Sci.* **86**, 844–874 (1960).
- [164] M. Hansen and W. Sickel, Best m -term approximation and tensor product of Sobolev and Besov spaces—the case of non-compact embeddings, *East J. Approx.* **16**(4) (2010), 345–388.
- [165] M. Hansen and W. Sickel, Best m -term approximation and Lizorkin-Triebel spaces, *J. Approx. Theory* **163**(8) (2011), 923–954.
- [166] M. Hansen and W. Sickel, Best m -term approximation and Sobolev-Besov spaces of dominating mixed smoothness—the case of compact embeddings, *Constr. Approx.* **36**(1) (2012), 1—51.

- [167] M. Hansen and J. Vybírál, The Jawerth-Franke embedding of spaces with dominating mixed smoothness, *Georgian Mathematical Journal*, **16**(4) (2009), 667–682.
- [168] S. Heinrich, Some open problems concerning the star-discrepancy, *J. Complexity* **19** (2003) 416–419.
- [169] S. Heinrich, E. Novak, G. Wasilkowski, H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension, *Acta Arith.* **96**(3) (2001) 279–302.
- [170] F. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter, Multi-level Monte Carlo algorithms for infinite-dimensional integration on \mathbb{R}^N , *J. Complexity*, **26**, (2010), 229–254.
- [171] F. Hickernell and X. Wang, The error bounds and tractability of quasi-Monte Carlo algorithms in infinite dimension, *Mathematics of Computation*, **71**, (2002), 1641–1661.
- [172] A. Hinrichs, Covering numbers, Vapnik–Cervonenkis classes and bounds for the star-discrepancy, *J. Complexity* **20** (2004), 477–483.
- [173] A. Hinrichs. Discrepancy of Hammersley points in Besov spaces of dominating mixed smoothness. *Math. Nachr.*, **283**(3):478–488, 2010.
- [174] A. Hinrichs, A. Kolleck, J. Vybírál, Carl’s inequality for quasi-Banach spaces, *J. Funct. Anal.* **271**(8), pp. 2293–2307, 2016.
- [175] A. Hinrichs, L. Markhasin, J. Oettershagen, T. Ullrich, Optimal quasi-Monte Carlo rules on order 2 digital nets for the numerical integration of multivariate periodic functions, *Num. Math.*, **134**:163–196, 2016.
- [176] A. Hinrichs, E. Novak, M. Ullrich, On weak tractability of the Clenshaw-Curtis Smolyak algorithm, *J. Approx. Theory* **183** (2014), 31–44.
- [177] A. Hinrichs, J. Oettershagen, Optimal point sets for quasi—Monte Carlo integration of bivariate periodic functions with bounded mixed derivatives, *Proceedings MCQMC 2014*, Leuven. arXiv e-print 1409.5894, 2014.
- [178] K. Höllig, Diameters of classes of smooth functions, in *Quantitative Approximation*, Academic Press, New York (1980), 163–176.
- [179] Hua Loo Keng and Wang Yuan, *Application of number theory to numerical analysis*, Springer-Verlag, Berlin, 1981.
- [180] R.S. Ismagilov, Widths of sets in normed linear spaces and the approximation of functions by trigonometric polynomials, *Uspekhi Mat. Nauk*, **29** (1974), 161–178; English transl. in *Russian Math. Surveys*, **29** (1974).
- [181] R. Ismagilov, Widths of sets in normed linear spaces and the approximation of functions by trigonometric polynomials, *UMN* 29(1974), 161–178; English transl. in *RMS* 29(1974).
- [182] A.D. Izaak, Kolmogorov widths in finite-dimensional spaces with mixed norm, *Math. Notes* **55**(1–2), pp- 30–36, 1994.
- [183] A.D. Izaak, Widths of Hölder-Nikolskii classes and finite-dimensional sets in spaces with a mixed norm, *Math. Notes* **59**(3–4), pp. 328–330, 1996.

- [184] B. Jawerth, Some observations on Besov and Lizorkin-Triebel spaces, *Math. Scand.* **40**(1) (1977), 94–104.
- [185] C. Kacwin, Realization of the Frolov cubature formula via orthogonal Chebychev-Frolov lattices, Master thesis, Institute for Numerical Simulation, University of Bonn, 2016.
- [186] C. Kacwin, J. Oettershagen, and T. Ullrich, On the orthogonality of the Chebyshev-Frolov lattice and applications. ArXiv e-prints, 2016. arXiv:1606.00492 [math.NA].
- [187] L. Kämmerer, Reconstructing hyperbolic cross trigonometric polynomials by sampling along rank-1 lattices, *SIAM J. Numer. Anal.* **51**, pp. 2773–2796, 2013.
- [188] L. Kämmerer and S. Kunis, On the stability of the hyperbolic cross discrete Fourier transform, *Numer. Math.* **117**(3), 581–600, 2011.
- [189] L. Kämmerer, S. Kunis, and D. Potts, Interpolation lattices for hyperbolic cross trigonometric polynomials, *J. Complexity* **28**(1), 76–92, 2012.
- [190] L. Kämmerer, D. Potts, T. Volkmer, Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling, *J. Complexity* **31**, pp. 543–576, 2015.
- [191] L. Kämmerer, D. Potts, T. Volkmer, Approximation of multivariate periodic functions by trigonometric polynomials based on sampling along rank-1 lattice with generating vector of Korobov form, *J. Complexity* **31**, pp. 424–456, 2015.
- [192] B.S. Kashin, Widths of certain finite-dimensional sets and classes of smooth functions, *Izv. AN SSSR*, **41** (1977), 334–351; English transl. in *Math. Izv.* **11** (1977).
- [193] B.S. Kashin, On certain properties of the space of trigonometric polynomials with the uniform norm, *Trudy Mat. Inst. Steklov*, **145** (1980), 111–116; English transl. in *Proc. Steklov Inst. Math.*, 1981, Issue 1.
- [194] B.S. Kashin, On approximation properties of complete orthonormal systems, *Trudy Mat. Inst. Steklov* **172** (1985), 187–191; English transl. in *Proc. Steklov Inst. Math.*, **3** (1987), 207–211.
- [195] B.S. Kashin and V.N. Temlyakov, On best m -term approximations and the entropy of sets in the space L^1 , *Math. Notes*, **56** (1994), 1137–1157.
- [196] B.S. Kashin and V.N. Temlyakov, Estimate of approximate characteristics for classes of functions with bounded mixed derivative, *Math. Notes*, **58** (1995), 1340–1342.
- [197] B.S. Kashin and V.N. Temlyakov, On a norm and related applications, *Matem. Zametki*, **64** (1998), 637–640; English translation in *Mathematical Notes*, **64** (1998), 551–554.
- [198] B.S. Kashin and V.N. Temlyakov, On a norm and approximation characteristics of classes of functions of several variables, *Metric theory of functions and related problems in analysis*, *Izd. Nauchno-Issled. Aktuarno-Finans. Tsentra (AFTs)*, Moscow, 1999, 69–99.
- [199] B.S. Kashin and V.N. Temlyakov, The volume estimates and their applications, *East J. Approx.*, **9** (2003), 469–485.

- [200] B.S. Kashin and V.N. Temlyakov, On a norm and approximate characteristics of classes of multivariate functions, *J. Math. Sciences*, **155** (2008), 57–80.
- [201] A.N. Kolmogorov, Über die beste Annäherung von Funktionen einer Funktionklasse, *Ann. Math.*, **37**(1936), 107–111.
- [202] N.M. Korobov, The approximate computation of multiple integrals, *Dokl. Akad. Nauk SSSR*, **124** (1959), 1207–1210.
- [203] N.M. Korobov, Number-theoretic methods in numerical analysis, Fizmatgis, Moscow, 1963.
- [204] N.M. Korobov, The trigonometric sums and their applications, Nauka, Moscow, 1989.
- [205] D. Krieg and E. Novak, A universal algorithm for multivariate integration, *Found. Comp. Math.*, (2016), DOI:10.1007/s10208-016-9307-y.
- [206] J. Kuelbs and W.V. Li, Metric entropy and the small ball problem for Gaussian measures, *J. Functional Analysis*, **116** (1993), 133–157.
- [207] T. Kühn. A lower estimate for entropy numbers. *J. Approx. Theory* 110(2001), pp. 120–124.
- [208] T. Kühn, S. Mayer, and T. Ullrich, Counting via entropy: new preasymptotics for the approximation numbers of Sobolev embeddings, *SIAM Journ. Num. Anal.*, to appear, ArXiv e-prints, 2015. arXiv:1505.00631 [math.NA].
- [209] T. Kühn, W. Sickel and T. Ullrich, Approximation numbers of Sobolev embeddings – sharp constants and tractability, *J. Complexity* **30** 95–116, 2014.
- [210] T. Kühn, W. Sickel and T. Ullrich, Approximation of mixed order Sobolev functions on the d -torus–Asymptotics, preasymptotics and d -dependence, *Constr. Approx.*, 42:353–398, 2015.
- [211] R.J. Kunsch, Bernstein numbers and lower bounds for the Monte Carlo error, *Proceedings MCQMC 2014*, Leuven.
- [212] F. Kuo, I. Sloan, G. Wasilkowski, and H. Woźniakowski, Liberating the dimension, *Journal of Complexity*, 26, (2010), 422–454.
- [213] F. Kuo, I. Sloan, and H. Woźniakowski, Lattice rules for multivariate approximation in the worst-case setting, In H. Niederreiter and D. Talay, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pp. 289–330, Springer Berlin Heidelberg, Berlin, 2006.
- [214] F. Kuo, G.W. Wasilkowski, and H. Woźniakowski, Lattice algorithms for multivariate approximation in the worst-case setting, *Constr. Approx.* **30**, pp. 475–493, 2009.
- [215] M. Levin, On the lower bound in the lattice point remainder problem for a parallelepiped, arxiv e-print 2013, arXiv:1307.2080.
- [216] M. Levin, On the lower bound of the discrepancy of Halton’s sequence, arxiv e-print 2014, arXiv:1412.8705.
- [217] M. Levin, On the lower bound of the discrepancy of $(t; s)$ sequences: I, arxiv e-print 2015, arXiv:1505.06610.

- [218] W. Li, W. Linde, Approximation, metric entropy and small ball estimates for Gaussian measures, *Ann. Probab.* **27**(3) (1999), 1556–1578.
- [219] M.A. Lifshits and B.S. Tsirelson, Small deviations of Gaussian fields, *Teor. Probab. Appl.*, **31** (1986), 557–558.
- [220] G. Lorentz, M. von Golitschek, Y. Makovoz. Constructive Approximation: Advanced Problems. Volume 304 of Grundlehren der Mathematischen Wissenschaften, Springer, Berlin(1996).
- [221] G.G. Magaril-Il’yaev, Intermediate derivatives, *Mat. Zametki* 25:1(1979), 81-96.
- [222] G.G. Magaril-Il’yaev, Inequalities of Bernstein-Nikol’skii type and approximation of generalized Sobolev classes, *Trudy Mat. Inst. Steklov.* 173(1986), 190-204.
- [223] G.G. Magaril-Il’yaev, On the approximations of Sobolev classes of functions on \mathbf{R}^n , *Trudy Mat. Inst. Steklov.* 180(1987), 154-155.
- [224] G.G. Magaril-Il’yaev, Trigonometric widths of Sobolev classes of functions on \mathbf{R}^n , *Trudy Mat. Inst. Steklov.* 181(1988), 147-155.
- [225] V.E. Maiorov, Discretization of the problem of diameters, *Uspekhi Mat. Nauk*, **30**(6)(186), 179–180, 1975.
- [226] V.E. Maiorov, On various widths of the class H_p^r in the space L_q , *Izv. Akad. Nauk SSSR Ser. Mat.*, **42** (1978), 773–788; English transl. in *Math. USSR-Izv.*, 1979, **13**.
- [227] V.E. Maiorov, Trigonometric diameters of the Sobolev classes W_p^r in the space L_q , *Math. Notes* **40** (1986), 590–597.
- [228] Y. Makovoz, On trigonometric n -widths and their generalizations, *J. Approx. Theory* **41** (1984), 361-366.
- [229] Yu.V. Malykhin, K.S. Ryutin, Product of octahedra is badly approximated in the $\ell_{2,1}$ -metric, arxiv e-print 2016, arXiv:1606.00738.
- [230] L. Markhasin, Discrepancy and integration in function spaces with dominating mixed smoothness, *Dissertationes Math.* **494**, 1–81, 2013.
- [231] P. Mathé, s -Numbers in information-based complexity, *J. Complexity*, **6**(1990), 41–66.
- [232] J. Matoušek, Geometric Discrepancy, Springer, Berlin, 1999.
- [233] B.S. Mityagin, Approximation of functions in the space L^p and C on the torus, *Mat. Sb.*, **58** (1962), 397–414.
- [234] N. Zh. Nauryzbaev and N. Temirgaliev, On the order of discrepancy of the Smolyak grid, *Mat. Zametki*, **85** (2009), 947–950.
- [235] N. Nauryzbayev and N. Temirgaliyev, An Exact Order of Discrepancy of the Smolyak Grid and Some General Conclusions in the Theory of Numerical Integration, *Found. Comput. Math.*, **12** (2012), 139–172.
- [236] V.K. Nguyen, Bernstein numbers of embeddings of isotropic and dominating mixed Besov spaces, *ArXiv e-prints* arXiv:1411.7246, 2014.

- [237] V.K. Nguyen and W. Sickel, Weyl numbers of embeddings of tensor product Besov spaces, *ArXiv e-prints* arXiv:1409.7565, 2014.
- [238] V.K. Nguyen, Weyl and Bernstein numbers of embeddings of Sobolev spaces with dominating mixed smoothness, *ArXiv e-prints*, arXiv:1506.00115.
- [239] V.K. Nguyen, M. Ullrich and T. Ullrich, Change of variable in spaces of mixed smoothness and numerical integration of multivariate functions on the unit cube, *ArXiv e-prints* arXiv:1511.02036, 2015.
- [240] H. Niederreiter, Point sets and sequences with small discrepancy, *Monatsh. Math.*, **104** (1987), 273–337.
- [241] N. Nikol'skaya, Approximation of differentiable functions of several variables by Fourier sums in the L_p -metric, *Sibirsk. Mat. Zh.* 15(1974), 395–412; English transl. in *Siberian Math. J.* 15(1974).
- [242] S.M. Nikol'skii, On boundary properties of differentiable functions of several variables, (Russian), *Dokl. Akad. Nauk SSSR* **146** (1962), 542–545.
- [243] S.M. Nikol'skii, On stable boundary values of differentiable functions of several variables, (Russian), *Mat. Sb.* **61** (1963), 224–252.
- [244] S.M. Nikol'skii, *Approximation of functions of several variables and embedding theorems*, Springer, 1975.
- [245] P.-A. Nitsche, Best N -term approximation spaces for tensor product wavelet bases, *Constr. Approx.* **24**(2006), 49–70.
- [246] B. Niu, F. Hickernell, T. Müller-Gronbach, and K. Ritter, Deterministic multilevel algorithms for infinite-dimensional integration on \mathbb{R}^N , *Journal of Complexity*, 27, (2011), 331–351.
- [247] E. Novak, Smolyak algorithm, In: *Encyclopedia of Math. Supplement II*, Kluwer, 2000.
- [248] E. Novak, Some results on the complexity of numerical integration, *Proc. MCQMC 2014 Leuven, Springer Proceedings in Mathematics and Statistics 2016*, 161–183, *ArXiv e-prints*, 2014, arXiv:1409.6714.v1 [math.NA].
- [249] E. Novak, K. Ritter, High-dimensional integration of smooth functions over cubes, *Numer. Math.* **75**(1), pp. 79–97, 1996.
- [250] E. Novak, K. Ritter, Simple cubature formulas with high polynomial exactness, *Constr. Approx.* **15**(4), pp. 499–522, 1999.
- [251] E. Novak, D. Rudolf, Tractability of the approximation of high-dimensional rank one tensors, *Constr. Approx.* **43**(1) (2016), 1–13.
- [252] E. Novak and H. Woźniakowski, *Tractability of multivariate problems. Volume I: Linear information*. EMS, Zürich, 2008.
- [253] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems, Volume I: Linear Information*, EMS Tracts in Mathematics, Vol. 6. Eur. Math. Soc. Publ. House, Zürich 2008.

- [254] E. Novak and H. Woźniakowski, Tractability of multivariate problems. Volume II: Standard information for functionals. EMS, Zürich, 2010.
- [255] E. Novak and H. Woźniakowski, Tractability of multivariate problems. Volume III: Standard information for operators. EMS, Zürich, 2012.
- [256] A. Pajor and N. Tomczak-Jaegermann, Subspaces of small codimension of finite-dimensional Banach spaces, *Proc. Amer. Math. Soc.*, **97**, 1986.
- [257] A. Papageorgiou and H. Woźniakowski, Tractability through increasing smoothness, *Journal of Complexity*, 26, (2010), 409-421.
- [258] A. Pietsch, Approximation spaces, *Journ. Approx. Theory* **32** (1980), 115–134.
- [259] A. Pietsch, History of Banach Spaces and Linear Operators, Birkhäuser, Basel, 2007.
- [260] A. Pietsch, Bad properties of the Bernstein numbers, *Studia Math.* **184** (2008), 263–269.
- [261] A. Pinkus. *N*-widths in Approximation Theory. Springer-Verlag, Berlin, 1985.
- [262] L. Plaskota and G. Wasilkowski, Tractability of infinite-dimensional integration in the worst case and randomized settings, *Journal of Complexity*, 27, (2011), 505.
- [263] L. Plaskota, G. Wasilkowski, and H. Woźniakowski, A new algorithm and worst case complexity for Feynman-Kac path integration, *J. Computational Physics*, 164, (2000), 335-353.
- [264] M.K. Potapov, B.V. Simonov and S.Yu. Tikhonov, Mixed moduli of smoothness in L_p , $1 < p < \infty$: a survey, *Surv. Approx. Theory* **8** (2013), 1–57.
- [265] N.N. Pustovoirov, On best approximations by analogs of “proper” and “improper” hyperbolic crosses, *Mat. Zametki*, **93** (2013), 466–476.
- [266] N.N. Pustovoirov, Multidimensional Jackson theorem in the L_2 space, *Mat. Zametki*, **49** (1991), 154–156.
- [267] J. Ratsaby and V. Maiorov, The degree of approximation of sets in Euclidean space using sets with bounded Vapnik-Chervonekis dimension, *Discrete Applied Math.* **86**(1998), 81–93.
- [268] J. Ratsaby and V. Maiorov, On the degree of approximation by manifolds of finite pseudo-dimension, *Constr. Approx.* **15**(1999), 291–300.
- [269] A. Romanyuk, Approximation of Besov classes of multivariate periodic functions in the space L_q , *Ukranian Math. J.* **43**(1991), 1398-1408.
- [270] A. Romanyuk, Linear widths of the Besov classes of periodic functions of many variables. I, *Ukranian Math. J.* **53**(2001), 744-760.
- [271] A.S. Romanyuk, Best M -term trigonometric approximations of Besov classes of periodic functions of several variables, *Izvestia RAN, Ser. Mat.* **67** (2003), 61–100; English transl. in *Izvestiya: Mathematics* (2003), 67(2):265–302.
- [272] A. Romanyuk, Linear widths of the Besov classes of periodic functions of many variables. II, *Ukr. Math. J.* **53**(2001), 965-977.

- [273] A. Romanyuk, Approximations of the classes $B_{p,\theta}^r$ of periodic functions of several variables by linear methods and best approximations, *Mat. Sb.*, **195**:2, 91–116, 2004.
- [274] A. Romanyuk, Kolmogorov widths of Besov classes $B_{p,\theta}^r$ in the metric of the space L_∞ , *Ukr. Mat. Visn.* **2**(2005), 201–218. English translation in *Ukr. Math. Bull.* **2** (2005), 205–222.
- [275] A. Romanyuk, Kolmogorov and trigonometric widths of Besov classes $B_{p,\theta}^r$ of multivariate periodic functions, *Mat. Sb.* 197(2006), 71–96.
- [276] A.S. Romanyuk, Bilinear and trigonometric approximations of periodic functions of several variables of Besov classes $B_{p,\theta}^r$, *Izvestia RAN, Ser. Math.*, **70** (2006), 69–98.
- [277] A. Romanyuk, Best approximations and widths of classes of periodic functions of several variables, *Mat. Sb.* 199(2008), 93–144.
- [278] A.S. Romanyuk, Approximation of classes $B_{p,\theta}^r$ of periodic functions of one and several variables, *Math. Notes* **87** (3) (2010), 403–415.
- [279] A.S. Romanyuk, Diameters and best approximation of the classes $B_{p,\theta}^r$ of periodic functions of several variables, *Anal. Math.*, **37** (2011), 181–213.
- [280] K.F. Roth, On irregularities of distribution, *Mathematica*, **1** (1954), 73–79.
- [281] K.F. Roth, On irregularities of distribution. II, *Communications on Pure and Applied Math.*, **29** (1976), 749–754.
- [282] K.F. Roth, On irregularities of distribution. III, *Acta Arith.*, **35** (1979), 373–384.
- [283] K.F. Roth, On irregularities of distribution. IV, *Acta Arith.*, **37** (1980), 67–75.
- [284] V.A. Rvachev, Compactly supported solutions of functional-differential equations and their applications, *Russian Math. Surveys* **45**, 87–120, 1990.
- [285] C. Schütt, Entropy numbers of diagonal operators between symmetric Banach spaces, *J. Approx. Theory*, **40** (1984), 121–128.
- [286] H.-J. Schmeisser and W. Sickel, Spaces of functions of mixed smoothness and their relations to approximation from hyperbolic crosses, *Journ. Appr. Theo.* **128** (2004), 115–150.
- [287] H.-J. Schmeisser and H. Triebel. *Topics in Fourier analysis and function spaces*. A Wiley-Interscience Publication. John Wiley & Sons Ltd., Chichester, 1987.
- [288] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. *Math. Ann.*, **63** (1907), 433–476.
- [289] W.M. Schmidt, Irregularities of distribution.VII, *Acta Arith.*, **21** (1972), 45–50.
- [290] W.M. Schmidt, Irregularities of distribution X, in *Number Theory and Algebra*, Acad.Press, New York, 1977, 311–329.
- [291] R. Schneider and A. Uschmajew, Approximation rates for the hierarchical tensor format in periodic Sobolev spaces, *J. Complexity*, **30**(2):56–71. 2014

- [292] Ch. Schwab and C. Gittelsohn, Sparse tensor discretizations high-dimensional parametric and stochastic PDEs, *Acta Numerica* 20, 291–467(2011).
- [293] C. Schwab, E. Süli, R.A. Todor, Sparse finite element approximation of high-dimensional transport-dominated diffusion problems. *ESAIM Math. Model. Numer. Anal.* 42(05), 777–819, 2008.
- [294] A. Seeger and T. Ullrich, Haar projection numbers and failure of unconditional convergence in Sobolev spaces, *Mathem. Zeitschrift*, to appear, doi:10.1007/s00209-016-1697-7, arXiv:1507.01211 [math.CA].
- [295] A. Seeger and T. Ullrich, Lower bounds for Haar projections: deterministic examples, *Constr. Approx.*, to appear, doi:10.1007/s00365-016-9353-3, arXiv:1511.01470 [math.CA].
- [296] W. Sickel, Approximation from sparse grids and function spaces of dominating mixed smoothness, *Banach Center Publ.* **72**, Inst. of Math., Polish Acad. of Sciences, Warszawa 2006, 271–283.
- [297] W. Sickel and F. Sprengel, Interpolation on sparse grids and Nikol’skii-Besov spaces of dominating mixed smoothness, *J. Comp. Anal. Appl.* **1** (1999), 263–288.
- [298] W. Sickel and T. Ullrich, The Smolyak algorithm, sampling on sparse grids and function spaces of dominating mixed smoothness, *East Journal on Approximations* **13**(4) (2007), 387–425.
- [299] W. Sickel, T. Ullrich, Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross, *Journ. Approx. Theor.* **161**(2009), 748–786.
- [300] W. Sickel, T. Ullrich, Spline Interpolation on sparse grids, *Applicable Analysis* 90(2011), 337–383.
- [301] M.B. Sirkov, N. Temirgaliev, On an Algorithm for Constructing Uniformly Distributed Korobov Grids, *Mat. Zametki*, **87** (2010), 948–950.
- [302] M.M. Skriganov, Constructions of uniform distributions in terms of geometry of numbers, *Algebra and Analysis*, **6** (1994), 200–230.
- [303] I. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *Journal of Complexity*, 14, (1998), 1–33.
- [304] S.A. Smolyak, The ϵ -entropy of the classes $E_s^{\alpha k}(B)$ and $W_s^\alpha(B)$ in the metric L_2 , *Dokl. Akad. Nauk SSSR*, **131** (1960), 30–33.
- [305] S.A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk* **148**(1963), 1042–1045.
- [306] I.M. Sobol, The distribution of points in a cube and the approximate evaluation of integrals, *Zh. Vychisl. Mat. Mat. Fiz.*, **7**(1967), 784–802.
- [307] S.A. Stasyuk, Best approximation of periodic functions of several variables from the classes $MB_{p,\theta}^\omega$, *Ukrainian Math. J.*, **64**(1)(2012), 156–161.

- [308] S.A. Stasyuk, Approximation by Fourier sums and Kolmogorov widths for classes $MB_{p,\theta}^\Omega$ of periodic functions of several variables, *Trudy Inst. Mat. i Mekh. UrO RAN*, **20**(1) (2014), 247–257.
- [309] S.A. Stasyuk, Best m -term trigonometric approximation for periodic multivariate functions with low mixed smoothness from Nikolskii–Besov-type classes, *Ukr. Mat. Zh.* **68** (7) (2016), 983–1003 (in Ukrainian).
- [310] S.B. Stechkin, On absolute convergence of orthogonal series, *Dokl. AN SSSR*, **102** (1955), pp. 37–40 (in Russian).
- [311] M. Stesin, Alexander widths of finite-dimensional sets and classes of smooth functions, *Dokl. 220*(1975), 1278–1281.
- [312] V.N. Sudakov, Gaussian random processes and measures of solid angles in Hilbert spaces, *Sov. Math. Dokl.*, **12** (1971), 412–415.
- [313] M. Talagrand, The small ball problem for the Brownian sheet, *Ann. Probab.*, **22** (1994), 1331–1354.
- [314] S.A. Telyakovskii, On estimates of the derivatives of trigonometric polynomials in several variables, *Sibirsk. Mat. Zh.*, **4** (1963), 1404–1411.
- [315] S.A. Telyakovskii, Some bounds for trigonometric series with quasi-convex coefficients, *Mat. Sb.* **63**(105) 1964, 426–444; English transl. in *Amer. Math. Soc. Transl.* (2) **86** (1970).
- [316] S.A. Telyakovskii, Work on the theory of approximation of functions carried out at the V.A. Steklov Institute of Mathematics, *Trudy Steklov Inst. Math.* **182** (1988), 128–180; English transl. in *Proceedings of the Steklov Institute of Mathematics*, **182** (1990), 141–197.
- [317] N. Temirgaliev, Application of the divisors theory to numerical integration of periodic functions in several variables, *Mat. Sb.*, **181** (1990), 490–505.
- [318] N. Temirgaliev, Efficiency of numerical integration algorithms related to divisor theory in cyclotomic fields, *Mat. Zametki*, **61** (1997), 297–301.
- [319] N. Temirgaliev, Tensor products of functionals and their application, *Doklady Mathematics*, **81** (2010), 78–82.
- [320] N. Temirgaliev, S.S. Kudaibergenov, and A.A. Shomanova, Applications of Smolyak quadrature formulas to the numerical integration of Fourier coefficients and in function recovery problems, *Russian Mathematics (Iz VUZ)* **54:3** (2010), 45–62.
- [321] N. Temirgaliev, N.Zh. Nauryzbayev, and A.A. Shomanova, Approximative capabilities of Smolyak type computational aggregates with Dirichlet, Fejér and Vallée-Poussin kernels in the scale of Ulyanov classes, *Russian Mathematics (Iz VUZ)* **59:7** (2015), 67–72.
- [322] N. Temirgaliev, S. S. Kudaibergenov, A. A. Shomanova, An application of tensor products of functionals in problems of numerical integration, *Izvestiya: Mathematics*, **73** (2009), 393–434.

- [323] V.N. Temlyakov, Approximation of periodic functions of several variables with bounded mixed derivative, Dokl. AN SSSR **248** (1979), 527–531; English transl. in SMDokl 20 (1979).
- [324] V.N. Temlyakov, On the approximation of periodic functions of several variables with bounded mixed difference, Dokl. AN SSSR, **253** (1980), 544–548; English transl. in SMDokl 22 (1980).
- [325] V.N. Temlyakov, Widths of some classes of functions of several variables, Dokl 267 (1982), 314–317; English transl. in SMDokl 26 (1982).
- [326] V.N. Temlyakov, Approximate recovery of periodic functions of several variables, Mat Sb 128(170) (1985), 256–268; English transl. in Mat Sb 56 (1987).
- [327] V.N. Temlyakov, On estimates of ε -entropy and widths of classes of functions with bounded mixed derivative or difference, Dokl 301 (1988), 288–291; English transl. in SMDokl 38 (1989).
- [328] V.N. Temlyakov, Approximation by elements of a finite-dimensional subspace of functions from various Sobolev or Nikol'skii spaces, MZ 43 (1988), 770–786; English transl. in MN 43 (1988).
- [329] V.N. Temlyakov, Approximation of periodic functions of several variables with bounded mixed difference, Mat. Sb., **133** (1980), 65–85; English transl. in Math. USSR Sbornik **41** (1982).
- [330] V.N. Temlyakov, Approximation of periodic functions of several variables with bounded mixed derivative, Trudy MIAN **156** (1980), 233–260; English transl. in Proceedings of Steklov Inst. Math., **2** (1983).
- [331] V.N. Temlyakov, Approximation of functions with a bounded mixed difference by trigonometric polynomials, and the widths of some classes of functions, Izvestia AN SSSR, Ser. Mat. **46** (1982), 171–186; English transl. in MIzv 20 (1983).
- [332] V.N. Temlyakov, Quadrature formulae and recovery of number-theoretical nets from nodal values for classes of functions with small degree of smoothness, Uspekhi Mat. Nauk **40(244)** (1985), 203–204; English transl. in Russian Math. Surveys, **40** (1985), 223–224.
- [333] V.N. Temlyakov, On reconstruction of multivariate periodic functions based on their values at the knots of number-theoretical nets, Anal. Math., **12** (1986), 287–305.
- [334] V.N. Temlyakov, Approximation of periodic functions of several variables by trigonometric polynomials, and widths of some classes of functions, Izv. AN SSSR, **49** (1985), 986–1030; English transl. in Math. Izv. **27** (1986).
- [335] V.N. Temlyakov, Approximate recovery of periodic functions of several variables, Mat. Sb., **128** (1985), 256–268; English transl. in Math. USSR Sbornik **56** (1987), 249–261.
- [336] V.N. Temlyakov, Approximation of periodic functions of several variables by bilinear forms, Izvestiya AN SSSR, **50** (1986), 137–155; English transl. in Math. USSR Izvestija, **28** (1987), 133–150.

- [337] V.N. Temlyakov, On widths of function classes, *Dokl. Akad. Nauk SSSR*, **294** (1987), 1069–1072; English transl. in *Soviet Math. Dokl.*, **35** (1987), 639–642.
- [338] V.N. Temlyakov, On approximation of functions of several variables by the trigonometric polynomials with harmonics from the hyperbolic crosses, *Ukr. Math. Journal*, **41** (1989), 518–524.
- [339] V.N. Temlyakov, Estimates of best bilinear approximations of periodic functions, *Trudy MIAN* **181** (1988), 250–267; English transl. in *Proceedings of Steklov Inst. of Math.* (1989), no. 4.
- [340] V.N. Temlyakov, Approximation of functions with bounded mixed derivative, *Trudy MIAN*, **178** (1986), 1–112. English transl. in *Proc. Steklov Inst. Math.*, **1** (1989).
- [341] V.N. Temlyakov, Estimates of the best bilinear approximations of functions of two variables and some of their applications, *Mat. Sb.*, **134** (1987), 93–107; English transl. in *Math. USSR-Sb* **62** (1989), 95–109.
- [342] V.N. Temlyakov, On estimates of ϵ -entropy and widths of classes of functions with bounded mixed derivative or difference, *Dokl. Akad. Nauk SSSR*, **301** (1988), 288–291; English transl. in *Soviet Math. Dokl.*, **38**, 84–87.
- [343] V.N. Temlyakov, Estimates of best bilinear approximations of periodic functions, *Proceedings of the Steklov Institute of Mathematics*, **4** (1989), 275–293.
- [344] V.N. Temlyakov, Estimates of the asymptotic characteristics of classes of functions with bounded mixed derivative or difference, *Trudy Matem. Inst. Steklov*, **189** (1989), 138–168; English transl. in *Proceedings of the Steklov Institute of Mathematics*, 1990, Issue 4, 161–197.
- [345] V.N. Temlyakov, On a way of obtaining lower estimates for the errors of quadrature formulas, *Matem. Sbornik*, **181** (1990), 1403–1413; English transl. in *Math. USSR Sbornik*, **71** (1992).
- [346] V.N. Temlyakov, On universal cubature formulas, *Dokl. Akad. Nauk SSSR*, **316** (1991); English transl. in *Soviet Math. Dokl.*, **43**, (1991), 39–42.
- [347] V.N. Temlyakov, Error estimates for the Fibonacci quadrature formulas for classes of functions with bounded mixed derivative, *Trudy Matem. Inst. Steklova*, **200** (1991), 327–335; English transl. in *Proc. Steklov Inst. Math.*, 1993, issue 2.
- [348] V.N. Temlyakov, Estimates of Best Bilinear Approximations of Functions and Approximation Numbers of Integral Operators, *Mat. Zametki*, **51** (1992), 125–134; English transl. in *Math. Notes* **51** (1992), 510–517.
- [349] V.N. Temlyakov, Universality of the Fibonacci Cubature Formulas, *Lecture Notes in Math.*, **1550** (1993), 178–184.
- [350] V.N. Temlyakov, Bilinear Approximation and Related Questions *Proc. Steklov Inst. of Math.*, **4** (1993), 245–265.
- [351] V.N. Temlyakov, On approximate recovery of functions with bounded mixed derivative, *J. Complexity*, **9** (1993), 41–59.

- [352] V.N. Temlyakov, *Approximation of periodic functions*, Nova Science Publishes, Inc., New York., 1993.
- [353] V.N. Temlyakov, On error estimates for cubature formulas, *Trudy Matem.Inst. Steklova*, **207** (1994), 326–338; English transl. in *Proc. Steklov Inst. Math.*, 1995, issue 6.
- [354] V.N. Temlyakov, An inequality for trigonometric polynomials and its application for estimating the entropy numbers, *J. Complexity*, **11** (1995), 293–307.
- [355] V.N. Temlyakov, Some inequalities for multivariate Haar polynomials, *East J. Approx.*, **1** (1995), 61–72.
- [356] V.N. Temlyakov, An inequality for trigonometric polynomials and its application for estimating the Kolmogorov widths, *East J. Approximations*, **2** (1996), 253–262.
- [357] V.N. Temlyakov, On two problems in the multivariate approximation, *East J. Approx.*, **4** (1998), 505–514.
- [358] V.N. Temlyakov, Nonlinear Kolmogorov’s widths, *Matem. Zametki*, **63** (1998), 891–902.
- [359] V.N. Temlyakov, Nonlinear m -term approximation with regard to the multivariate Haar system, *East J. Approx.*, **4** (1998), 87–106.
- [360] V.N. Temlyakov, Greedy Algorithms with Regard to Multivariate Systems with Special Structure, *Constr. Approx.* **16** (2000), 399–425.
- [361] V.N. Temlyakov, The best m -term approximation and greedy algorithms, *Adv. Comput. Math.*, **8** (2001), 249–265.
- [362] V.N. Temlyakov, Greedy algorithms in Banach spaces, *Adv. Comput. Math.*, **14** (2001), 277–292.
- [363] V.N. Temlyakov, Universal bases and Greedy Algorithms for Anisotropic Function Classes, *Constructive Approximation* **18** (2002), 529–550.
- [364] V.N. Temlyakov, Nonlinear methods of approximation, *Found. Comput. Math.*, **3** (2003), 33–107.
- [365] V.N. Temlyakov, Cubature formulas, discrepancy, and nonlinear approximation, *J. Complexity*, **19** (2003), 352–391.
- [366] V.N. Temlyakov, Greedy-Type Approximation in Banach Spaces and Applications, *Constr. Approx.* **21** (2005), 257–292.
- [367] V.N. Temlyakov, On universal estimators in learning theory, *Trudy MIAN im. V.A. Steklova*, **255** (2006), 256–272; English transl. in *Proceedings of the Steklov Inst. Math.*, **255** (2006), 244–259.
- [368] V.N. Temlyakov, *Greedy approximation*, Cambridge University Press, 2011.
- [369] V.N. Temlyakov, An inequality for the entropy numbers and its application, *J. Approx. Theory*, **173** (2013), 110–121.
- [370] V.N. Temlyakov, Sparse approximation and recovery by greedy algorithms in Banach spaces, *Forum of Mathematics, Sigma*, **2** (2014), e12, 26 pages; IMI Preprint, 2013:09, 1–27; arXiv:1303.6811v1, 27 Mar 2013.

- [371] V.N. Temlyakov, Constructive sparse trigonometric approximation and other problems for functions with mixed smoothness, *Matem. Sb.*, **206** (2015), 131–160; see also arXiv:1412.8647v1 [math.NA] 24 Dec 2014.
- [372] V.N. Temlyakov, Constructive sparse trigonometric approximation for functions with small mixed smoothness, *Constr. Approx.*, to appear; arXiv:1503.00282v1 [math.NA] 1 Mar 2015.
- [373] V.N. Temlyakov, Incremental Greedy algorithm and its applications in numerical integration, Springer Proceedings in Mathematics and Statistics, Monte Carlo and Quasi-Monte Carlo Methods, MCQMC, Leuven, Belgium, April 2014, **163** (2016), 557–570.
- [374] V.N. Temlyakov, On the entropy numbers of the mixed smoothness function classes, arXiv:1602.08712v1 [math.NA] 28 Feb 2016.
- [375] V.N. Temlyakov and S. Tikhonov, Remez-type inequalities for the hyperbolic cross polynomials, arXiv:1606.03773v1 [math.CA], 12 June 2016.
- [376] V.M. Tikhomirov, Widths of sets in function spaces and the theory of best approximations, *Uspekhi Mat. Nauk*, 15(3), (93), (1960), 81–120. English translation in Russian Math. Survey, 15, 1960.
- [377] V.M. Tikhomirov, Some topics in approximation theory, Moscow State Univ., Moscow, 1976.
- [378] V.M. Tikhomirov, Approximation theory, *Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat. Fund. Napr.*, 1987, Volume 14, 103–260.
- [379] V.M. Tikhomirov, Analysis II: Convex Analysis and Approximation Theory (Eds. Gamkrelidze) Springer, 1990.
- [380] M.F. Timan, Imbedding classes of functions in $L_p^{(k)}$, *Izv. Vyssh. Uchebn. Zaved., Mat.*, **10**, 61–74, 1974.
- [381] H. Tran, C.G. Webster, G. Zhang, Analysis of quasi-optimal polynomial approximations for parameterized PDEs with deterministic and stochastic coefficients, *Arxiv e-prints*, arXiv:1508.01821v1 [math.AP].
- [382] H. Triebel, *Bases in function spaces, sampling, discrepancy, numerical integration*, European Math. Soc. Publishing House, Zürich, 2010.
- [383] H. Triebel. *Faber systems and their use in sampling, discrepancy, numerical integration*. EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2012.
- [384] H. Triebel, Global solutions of Navier–Stokes equations for large initial data belonging to spaces with dominating mixed smoothness, *J. Complexity* 31, 147–161, 2015.
- [385] R.M. Trigub and E.S. Belinsky, Fourier Analysis and Approximation of Functions, Kluwer Academic Publishers, 2004.
- [386] M. Ullrich, On "Upper error bounds for quadrature formulas on function classes" by K. K. Frolov, Proc. MCQMC 2014 Leuven, Springer Proceedings in Mathematics and Statistics 2016, 571–582, *ArXiv e-prints* arXiv:1404.5457, 2014.

- [387] M. Ullrich, A Monte Carlo method for integration of multivariate smooth functions, *SIAM J. Numer. Anal.*, to appear, *ArXiv e-prints* arXiv:1604.06008, 2016.
- [388] M. Ullrich and T. Ullrich, The role of Frolov’s cubature formula for functions with bounded mixed derivative, *SIAM Journ. Num. Anal.*, **54**(2), pp. 969–993; ArXiv e-prints, 2015. arXiv:1503.08846 [math.NA].
- [389] T. Ullrich, Über die periodische Interpolation auf schwach besetzten Gittern mittels de la Vallée Poussin-Kernen, Diploma thesis, Jena, 2004.
- [390] T. Ullrich, Function spaces with dominating mixed smoothness; characterization by differences. *Jenaer Schriften zur Mathematik und Informatik*, Math/Inf/05/06, 2006.
- [391] T. Ullrich, Smolyak’s algorithm, sampling on sparse grids and Sobolev spaces of dominating mixed smoothness, *East Journal on Approximations*, **14**(1):1–38, 2008.
- [392] T. Ullrich, Optimal cubature in Besov spaces with dominating mixed smoothness on the unit square, *J. Complexity* **30**, 72–94, 2014.
- [393] P.L. Ul’yanov, Imbedding theorems and relations between best approximations (moduli of continuity) in different metrics, *Mat. Sb. (N.S.)*, **81**(123), No. 1, 104–131, 1970.
- [394] I.V. Vilenkin, Plane nets of integration, *Zhur. Vychisl. Mat. i Mat. Fis.*, **7** (1967), 189–196; English transl. in, *USSR Comp. Math. and Math. Phys.*, **7** (1967), 258–267.
- [395] S. M. Voronin and N. Temirgaliev, Quadrature formulas associated with divisors of the field of Gaussian numbers, *Mat. Zametki*, **46** (1989), 34–41.
- [396] J. Vybíral, Function spaces with dominating mixed smoothness, *Dissertationes Math.* **436**, 73 pp., 2006.
- [397] J. Vybíral, A new proof of the Jawerth-Franke embedding, *Rev. Mat. Complut.* **21**(1), 75–82, 2008.
- [398] J. Vybíral, Widths of embeddings in function spaces, *J. Complexity*, **24**(4), pp. 545–570, 2008.
- [399] G. Wasilkowski, Liberating the dimension for function approximation and integration, in *Monte Carlo and Quasi-Monte Carlo Methods 2012*, L. Plaskota and H. Woźniakowski, eds., Springer, Heidelberg, (2012), 211–231.
- [400] G. Wasilkowski, Liberating the dimension for L_2 -approximation, *Journal of Complexity*, **28**, (2012), 304–319.
- [401] G. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, *J. of Complexity* **11**, 1–56, 1995.
- [402] G. Wasilkowski and H. Woźniakowski, On tractability of path integration, *J. Math. Physics*, **37** (1996), 2071–2088.
- [403] G. Wasilkowski and H. Woźniakowski, Liberating the dimension for function approximation, *Journal of Complexity*, **27**, (2011), 86–110.
- [404] G. Wasilkowski and H. Woźniakowski, Liberating the dimension for function approximation: Standard information, *Journal of Complexity*, **27**, (2011), 417–440.

- [405] H. Woźniakowski, Average case complexity of linear multivariate problems, *J. Complexity* **8** (1992), 337–372.
- [406] H. Yserentant, *Regularity and Approximability of Electronic Wave Functions*. Lecture Notes in Mathematics, Springer (2010).
- [407] C. Zenger, Sparse grids, in *Parallel Algorithms for Partial Differential Equations* (W. Hackbusch, ed.), Vol. 31 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden, 1991.
- [408] A. Zh. Zhubanysheva, N. Temirgaliev, Zh. N. Temirgalieva, Application of divisor theory to the construction of tables of optimal coefficients for quadrature formulas, *Zh. Vychisl. Mat. Fiz.*, **49** (2009), 14–25.